

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 12:40:24 ; Search time 71 Seconds
(without alignments)
26.999 Million cell updates/sec

Title: US-09-806-302A-2
Perfect score: 496
Sequence: 1 MKLMLVLAALLHCRVADS.....NFGIMHTYDSIMCNKSN 95

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

1: /cgn2-6/ptodata/2/pubpaa/US08_NEM_PUB.pep:*
2: /cgn2-6/ptodata/2/pubpaa/PCIT_NEM_PUB.pep:*
3: /cgn2-6/ptodata/2/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2-6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2-6/ptodata/2/pubpaa/US07_NEM_PUB.pep:*
6: /cgn2-6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2-6/ptodata/2/pubpaa/PCITUS_PUBCOMB.pep:*
8: /cgn2-6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2-6/ptodata/2/pubpaa/US09_NEM_PUB.pep:*
10: /cgn2-6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2-6/ptodata/2/pubpaa/US10_NEM_PUB.pep:*
12: /cgn2-6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2-6/ptodata/2/pubpaa/US60_NEM_PUB.pep:*
14: /cgn2-6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496	100.0	95	10	US-09-110-716-31
2	496	100.0	95	10	US-09-985-911-6
3	408	82.3	77	10	US-09-110-716-13
4	393.5	79.3	76	10	US-09-110-716-40
5	281	56.7	93	10	US-09-757-417-27
6	281	56.7	93	10	US-09-934-054-3
7	281	56.7	93	10	US-09-934-054-10
8	281	56.7	93	12	US-10-007-805-503
9	281	56.7	132	10	US-09-757-417-47
10	281	56.7	410	12	US-10-007-805-495
11	281	56.7	743	12	US-10-007-805-494
12	281	56.7	1095	12	US-10-007-805-493
13	206	41.5	75	10	US-09-110-716-41
14	196.5	39.6	95	10	US-09-985-911-27
15	195.5	39.4	95	10	US-09-934-054-12
16	166	33.5	76	10	US-09-110-716-39
17	77	15.5	21	10	US-09-757-417-3
18	77	15.5	21	10	US-09-757-417-16
19	77	15.5	21	10	US-09-757-417-18

20	77	15.5	21	12	US-10-007-805-498	Sequence 498, App
21	72	14.5	20	10	US-09-757-417-32	Sequence 32, App1
22	68	13.7	55	10	US-09-757-417-28	Sequence 28, App1
23	67.5	13.6	112	10	US-09-934-054-8	Sequence 8, App1
24	67	13.5	20	10	US-09-757-417-30	Sequence 30, App1
25	66	13.3	20	10	US-09-757-417-4	Sequence 4, App1
26	66	13.3	20	10	US-10-007-805-493	Sequence 33, App1
27	66	13.3	20	12	US-10-007-805-499	Sequence 499, App
28	65.5	13.2	142	10	US-09-821-839-7	Sequence 7, App1
29	62.5	12.6	91	10	US-09-934-054-9	Sequence 6, App1
30	62.5	12.6	143	10	US-09-821-839-6	Sequence 9, App1
31	62.5	12.6	293	9	US-09-902-525-42	Sequence 42, App1
32	61	12.3	765	10	US-09-975-326-4	Sequence 4, App1
33	61	12.3	766	9	US-09-934-056-2	Sequence 2, App1
34	61	12.3	766	10	US-09-975-326-2	Sequence 2, App1
35	60.5	12.2	180	9	US-09-992-598-256	Sequence 256, App
36	60.5	12.2	180	9	US-09-989-293A-256	Sequence 256, App
37	60.5	12.2	180	9	US-09-989-735-256	Sequence 256, App
38	60.5	12.2	180	9	US-09-980-444-256	Sequence 256, App
39	60.5	12.2	180	9	US-09-989-730-256	Sequence 256, App
40	60.5	12.2	180	9	US-09-980-436-256	Sequence 256, App
41	60.5	12.2	180	9	US-09-991-181-256	Sequence 256, App
42	60.5	12.2	180	9	US-09-993-687-256	Sequence 256, App
43	60.5	12.2	180	9	US-09-989-734-256	Sequence 256, App
44	60.5	12.2	180	9	US-09-997-653-256	Sequence 256, App
45	60.5	12.2	180	9	US-10-174-590-196	Sequence 196, App

ALIGNMENTS

```
RESULT 1
US-09-110-716-31
: Sequence 31, Application US/09110716A
: Patent No. US20020034739A1
: GENERAL INFORMATION:
: APPLICANT: Lehner, Robert I.
: APPLICANT: Zhao, Chengquan
: APPLICANT: Glasgow, Benjamin J.
: TITLE OF INVENTION: PEPTIDE CHARACTERISTIC OF CERTAIN TUMORS
: FILE REFERENCE: 22000-20596.00
: CURRENT APPLICATION NUMBER: US/09/110,716A
: CURRENT FILING DATE: 1998-07-07
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 31
: LENGTH: 95
: TYPE: PRT
: ORGANISM: lipophilin C
US-09-110-716-31

Query Match      100.0%; Score 496; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.5e-48;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLMLVLAALLHCRVADSCSKLLEDMVEKXTINSISIPYKELDFIDSDAAEAMG 60
Db 1 MKLMLVLAALLHCRVADSCSKLLEDMVEKXTINSISIPYKELDFIDSDAAEAMG 60
QY 61 KFKOCFLNQSHTLKNFGIMHTYDSIMCNKSN 95
Db 61 KFKOCFLNQSHTLKNFGIMHTYDSIMCNKSN 95

RESULT 2
US-09-985-911-6
: Sequence 6, Application US/09985911
: Patent No. US20020151012A1
: GENERAL INFORMATION:
: APPLICANT: NI ET AL.
: TITLE OF INVENTION: HUMAN ENDOMETRIAL SPECIFIC STEROID-BINDING FACTOR I, II AND II
: FILE REFERENCE: PF257D3
: CURRENT APPLICATION NUMBER: US/09/985,911
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 12:40:24 ; Search time 71 Seconds
(without alignments)
26.999 Million cell updates/sec

Title: US-09-806-302a-2
Perfect score: 496
Sequence: 1 MKLWMLMALILHCYADS.....NFGIMHTVYDSIMCNKSN 95

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496	100.0	95	10	US-09-110-716-31
2	496	100.0	95	10	US-09-985-911-6
3	408	82.3	77	10	US-09-110-716-13
4	393.5	79.3	76	10	US-09-110-716-40
5	281	56.7	93	10	US-09-757-417-27
6	281	56.7	93	10	US-09-934-054-3
7	281	56.7	93	10	US-09-934-054-10
8	281	56.7	93	12	US-10-007-805-503
9	281	56.7	132	10	US-09-757-417-471
10	281	56.7	410	12	US-10-007-805-495
11	281	56.7	743	12	US-10-007-805-494
12	281	56.7	1095	12	US-10-007-805-493
13	206	41.5	75	10	US-09-110-716-41
14	196.5	39.6	95	10	US-09-985-911-27
15	195.5	39.4	95	10	US-09-934-054-12
16	166	33.5	76	10	US-09-110-716-39
17	77	15.5	21	10	US-09-757-417-3
18	77	15.5	21	10	US-09-757-417-16
19	77	15.5	21	10	US-09-757-417-18

20	77	15.5	21	12	US-10-007-805-498	Sequence 498, App
21	72	14.5	20	10	US-09-757-417-32	Sequence 32, Appl
22	68	13.7	55	10	US-09-757-417-28	Sequence 28, Appl
23	67.5	13.6	112	10	US-09-934-054-8	Sequence 8, Appl
24	67	13.3	20	10	US-09-757-417-30	Sequence 30, Appl
25	66	13.3	20	10	US-09-757-417-4	Sequence 4, Appl
26	66	13.3	20	10	US-09-757-417-33	Sequence 33, Appl
27	66	13.3	20	12	US-10-007-805-499	Sequence 499, App
28	65.5	13.2	142	10	US-09-821-839-7	Sequence 7, Appl
29	62.5	12.6	91	10	US-09-934-054-9	Sequence 9, Appl
30	62.5	12.6	143	10	US-09-821-839-6	Sequence 6, Appl
31	62.5	12.6	293	10	US-09-902-525-42	Sequence 42, Appl
32	61	12.3	765	10	US-09-975-326-4	Sequence 4, Appl
33	61	12.3	766	9	US-09-934-406-2	Sequence 2, Appl
34	61	12.3	766	10	US-09-975-326-2	Sequence 2, Appl
35	60.5	12.2	180	9	US-09-992-598-256	Sequence 256, App
36	60.5	12.2	180	9	US-09-989-293A-256	Sequence 256, App
37	60.5	12.2	180	9	US-09-989-735-256	Sequence 256, App
38	60.5	12.2	180	9	US-09-990-444-256	Sequence 256, App
39	60.5	12.2	180	9	US-09-989-730-256	Sequence 256, App
40	60.5	12.2	180	9	US-09-990-436-256	Sequence 256, App
41	60.5	12.2	180	9	US-09-991-181-256	Sequence 256, App
42	60.5	12.2	180	9	US-09-993-687-256	Sequence 256, App
43	60.5	12.2	180	9	US-09-989-734-256	Sequence 256, App
44	60.5	12.2	180	9	US-09-997-653-256	Sequence 256, App
45	60.5	12.2	180	9	US-10-174-590-196	Sequence 196, App

ALIGNMENTS

RESULT 1
US-09-110-716-31
; Sequence 31, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596.00
; CURRENT APPLICATION NUMBER: US/09/110, 716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 31
; LENGTH: 95
; TYPE: PRT
; ORGANISM: lipophilin C
US-09-110-716-31

Query Match 100.0%; Score 496; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.5e-48;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWMLMALILHCYADSGCKLEPMVEKTNDSISPEYKELQEFIDSDAAFAWG 60
DB 1 MKLWMLMALILHCYADSGCKLEPMVEKTNDSISPEYKELQEFIDSDAAFAWG 60
QY 61 KFKCFLNOSHRTLKNFGGLMMHTVYDSIMCNKSN 95
DB 61 KFKCFLNOSHRTLKNFGGLMMHTVYDSIMCNKSN 95

RESULT 2
US-09-985-911-6
; Sequence 6, Application US/0985911
; Patent No. US20020151012A1
; GENERAL INFORMATION:
; APPLICANT: NI ET AL.
; TITLE OF INVENTION: HUMAN ENDOMETRIAL SPECIFIC STEROID-BINDING FACTOR I, II AND III
; FILE REFERENCE: PF257D3
; CURRENT APPLICATION NUMBER: US/09/985, 911

```
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/583,169
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 09/263,810
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 08/921,451
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: 60/014,724
; PRIOR FILING DATE: 1996-03-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 95
; TYPE: PRT
; ORGANISM: human
US-09-985-911-6

Query Match          100.0%; Score 496; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 4,5e-48;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLMVLMLAALLHCVADSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMG 60
DB 1 MRLMVLMLAALLHCVADSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMG 60
DB 61 KFKQCFLNQSHRTLKNFGLMHTVYDSIWCNMKSN 95
DB 61 KFKQCFLNQSHRTLKNFGLMHTVYDSIWCNMKSN 95

RESULT 3
US-09-110-716-13
; Sequence 13, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596.00
; CURRENT APPLICATION NUMBER: US/09/110,716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-110-716-13

Query Match          82.3%; Score 408; DB 10; Length 77;
Best Local Similarity 100.0%; Pred. No. 1,8e-38;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 DSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHRTLKNFG 78
DB 1 DSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHRTLKNFG 60
QY 79 LMHTVYDSIWCNMKSN 95
DB 61 LMHTVYDSIWCNMKSN 77

RESULT 4
US-09-110-716-40
; Sequence 40, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596.00
```

```
; CURRENT APPLICATION NUMBER: US/09/110,716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Lpmc
US-09-110-716-40

Query Match          79.3%; Score 393.5; DB 10; Length 76;
Best Local Similarity 98.7%; Pred. No. 7.1e-37;
Matches 76; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 19 DSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHRTLKNFG 78
DB 1 DSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMGKFKQCFLNQSHRTLKNFG 60
QY 79 LMHTVYDSIWCNMKSN 95
DB 61 LMHTVYDSIWCNMKSN 76

RESULT 5
US-09-757-417-27
; Sequence 27, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-417-27

Query Match          56.7%; Score 281; DB 10; Length 93;
Best Local Similarity 58.7%; Pred. No. 2.3e-24;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 MRLMVLMLAALLHCVADSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMG 60
DB 1 MRLMVLMLAALLHCVADSGCKLLEDVVEKTSIPYKELQEFIDSDAAAEAMG 60
DB 61 KFKQCFLNQSHRTLKNFGLMHTVYDSIWCNM 92
DB 61 KFKQCFLNQSHRTLKNFGLMHTVYDSIWCNM 92

RESULT 6
US-09-934-054-3
; Sequence 3, Application US/09934054
; Patent No. US20020107385A1
; GENERAL INFORMATION:
; APPLICANT: Akerbloom, Ingrid E.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Murry, Lynn E.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 2200 Sand Hill Road, Suite 100
```

```

1      CITY: Menlo Park
2      STATE: CA
3      COUNTRY: USA
4      ZIP: 94025-6936
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: Floppy disk
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: Patent Release #1.0, Version #1.25
11
12     CURRENT APPLICATION DATA:
13     APPLICATION NUMBER: US/09/934,054
14     FILING DATE: 21-Aug-2001
15     PRIOR APPLICATION DATA:
16     APPLICATION NUMBER: US/08/747,547
17     FILING DATE: <unknown>
18
19     ATTORNEY/AGENT INFORMATION:
20     NAME: Billings, Lucy J.
21     REGISTRATION NUMBER: 36,749
22     REFERENCE/DOCKET NUMBER: PF-0077 US
23     TELECOMMUNICATION INFORMATION:
24     TELEPHONE: (650) 855-0555
25     TELEFAX: (650) 845-4166
26
27     INFORMATION FOR SEQ ID NO: 3:
28     SEQUENCE CHARACTERISTICS:
29     LENGTH: 93 amino acids
30     TYPE: amino acid
31     STRANDEDNESS: double
32     TOPOLOGY: linear
33
34     MOLECULE TYPE: <unknown>
35
36     SEQUENCE DESCRIPTION: SEQ ID NO: 3:
37     US-09-934-054-3

```

Query Match 56.7%; Score 281; DB 10; Length 93;
Best Local Similarity 58.7%; Pred. No. 2.3e-24;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLVMTLALLLHCYADSGCKLEDDWERTINSDISIPETKELLQEFIDSDAAAEANG 60
||||| | ||||| : ||| :
Db 1 MKLLVMTLALALSHCGYAGGCGPLEENISTKNPQVSKTEKKELQEFIDDNATTNMAID 60
||||| | ||||| : ||| :
QY 61 KFKQCFLNGSHRTLKNEFGLMHTVVDSIWCNA 92
: ||||| : : ||| :
Db 61 ELKECFNLQDETLSNVEFMQLIVDSEICDL 92

RESULT 7
US-09-934-054-10
Sequence 10, Application US/09934054
Patent No. US20020107385A1
GENERAL INFORMATION:
APPLICANT: Akerblom, Ingrid E.
Hillman, Jennifer L.
Murphy, Lynn E.
Coff, Surya K.
Hawkins, Phillip R.
TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025-6936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/934,054
FILING DATE: 21-Aug-2001
PRIOR APPLICATION DATA:

```

1 APPLICATION NUMBER: US/08/747,547
2 FILING DATE: <Unknown>
3
4 ATTORNEY/AGENT INFORMATION:
5
6 NAME: Billings, Lucy J.
7 REGISTRATION NUMBER: 36,749
8 REFERENCE/DOCKET NUMBER: PE-0077 US
9
10 TELECOMMUNICATION INFORMATION:
11
12 TELEPHONE: (650) 855-0555
13
14 TELEFAX: (650) 845-4166
15
16 INFORMATION FOR SEQ ID NO: 10:
17
18 SEQUENCE CHARACTERISTICS:
19
20 LENGTH: 93 amino acids
21 TYPE: amino acid
22 STRANDEDNESS: double
23 TOPOLOGY: linear
24
25 MOLECULE TYPE: <Unknown>
26
27 SEQUENCE DESCRIPTION: SEQ ID NO: 10:
28
29 US-09-934-054-10

```

56	Query Match	56.7%	Score 281;	DB 10;	Length 93;
	Best Local Similarity	58.7%	Pred. No. 2,3e-24;		
Matches	54;	Conservative	13;	Mismatches	25; Indels 0; Gaps 0;
QY	1	MKLLMVLMLALLHLCVADSGCKLEDMVAVETINDSIDIPKPKELLOSFDSDAAALAMG	60		
Db	1	MKLLMVLMLALLHLCVADSGCKLEDMVAVETINDSIDIPKPKELLOSFDSDAAALAMG	60		
QY	61	KFKQCFNLQSHRTLKNFGIMAHVYDSTWCNN	92		
		:-:	:-:	:-:	:-:
Db	61	ELKECFNLQDTELNSNVEFMQLIYDSSLCIDL	92		
		:-:	:-:	:-:	:-:

```

RESULT 8
US-10-007-805-503
; Sequence 503, Application US/10007805
; Patent No. US20020150581A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
APPLICANT: Harlocker, Susan L.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: McNeill, Patricia D.
APPLICANT: Durham, Margarita
TITLE OR INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OR INVENTION: AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.470C10
CURRENT APPLICATION NUMBER: US/10/007,805
CURRENT FILING DATE: 2001-12-07
NUMBER OF SEQ ID NOS: 593
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 503
; LENGTH: 93
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-007-805-503

```

	Query Match	56.7%;	Score 281;	DB 12;	Length 93;
	Best Local Similarity	58.7%;	Pred. No. 2,3e-24;		
	Matches	54;	Conservative 13;	Mismatches 25;	Indels 0; Gaps 0;
QY	1	MKLILVMTALAILLCYADSGCKLEEDWVEKTIINSDISIPFYKELLOEFTIDSDAAAFAMG	60		
		: :			
Dd	1	MKLILVMTALAASQHCYAGSGCPLEENVIYSKTINPQVSKEYTEKKELLOEFTIDNATTNNAID	60		
QY	61	KFKCFGLNSHRTLKNFGIMHTYDVSITMCNN	92		
		: : : : : : : : : : : : : : : : : : : : : : :			
Dd	61	EKECEFLNQTDFTLSNVFEMQLYDSSISLCD	92		

Qy 1 MKLWVLMALALLHCYAD-SGCKLEDMVEKRTNSDISIPEYKELLOEFIDSDAAEAM 59
| | : : | : : | | : : : | | : : : | | : : : | :
Db 1 MKLVFLFLVLTIPICGYASGCSILDEVTRGTINSTVTILHDYMKLVKPYVDHFTKAV 60
Qy 60 GKFKQCFLNOSHRKTLNFGIMHTVYDSINCNKS 94
: | | | | : : : | : : : | : : : | :
Db 61 KQFKQCFLDQTDKTLNVGVMEAFNSESQOOPS 95

Search completed: January 25, 2003, 13:27:38
Job time : 72 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:19:09 ; Search time 82 seconds

(without alignments)
48.052 Million cell updates/sec

Title: US-09-806-302a-2

Sequence: 1 MKLLVLMALALHLHCYADS.....NFGLMHFRYYDSIMCMKSN 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496	100.0	95	1 MGBB_HUMAN	075556 homo sapien
2	281	56.7	93	1 MGBB_HUMAN	Q13396 homo sapien
3	195.5	39.4	95	1 PSC3_RAT	P02780 rattus norv
4	71.5	14.4	131	1 YRS9_MYCTG	03301 mycobacteri
5	70	14.1	303	1 FTSX_RICPR	005948 rickettsia
6	67.5	13.6	112	1 PSC2_RAT	P02781 rattus norv
7	67.5	13.6	268	1 TRPC_ACICA	P00911 acinetobact
8	66.5	13.4	234	1 RADC_HAEIN	P44952 haemophilus
9	65.5	13.2	162	1 IL15_MOUSE	P48346 mus musculu
10	65.5	13.2	603	1 VE1_HPV08	P06420 human papil
11	65	13.1	2144	1 GLT1_YEAST	Q12680 saccharomyc
12	64.5	13.0	153	1 Y156_UREPA	099498 ureaplasma
13	64.5	13.0	603	1 VE1_HPV21	P50759 human papil
14	64.5	13.0	611	1 DYNL_ARATH	P42697 arabidopsis
15	63.5	12.8	582	1 DOT1_YEAST	Q04089 saccharomyc
16	63.5	12.8	1294	1 YAB3_SCHPO	009716 schizosacch
17	63	12.7	1022	1 SC44_RICPR	092449 rickettsia
18	62.5	12.6	91	1 UTER_RABIT	P02779 oryctolagus
19	62.5	12.6	113	1 PTH_CALSI	P55321 callinectes
20	62.5	12.6	223	1 YW24_YEAST	P40218 saccharomyc
21	62.5	12.6	663	1 TAZ1_SCHPO	P79905 schizosacch
22	62	12.5	681	1 CA02_RABIT	002767 o acyl-coen
23	62	12.5	938	1 V120_HSVJ7	P52438 human herpe
24	62	12.5	1071	1 PR16_YEAST	P15938 saccharomyc
25	61.5	12.4	105	1 YR03_LISMO	Q8936 listeria mo
26	61.5	12.4	570	1 HEM1_KLUFA	P78698 kluyveromyc
27	61.5	12.4	734	1 METE_THEMA	Q94112 thermotoga
28	61.5	12.4	2190	1 CCAD_CHICK	073700 gallus gall
29	61	12.3	63	1 YE8B_ECOLI	P76275 escherichia
30	61	12.3	605	1 VE1_HPV14	P36721 human papil
31	61	12.3	606	1 VE1_HPV5B	P26542 human papil
32	61	12.3	732	1 K086_MOUSE	P27641 mus musculu
33	61	12.3	830	1 J1P2_MOUSE	Q9499 mus musculu

ALIGNMENTS

RESULT 1	ID	MGBB_HUMAN	STANDARD:	PRT:	95 AA.
AC	075556:				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Mammaglobin B precursor (Mammaglobin 2) (Lipophilin C) (Lacryglobin)				
DE	(Secretoglobin family 2A member 1).				
GN	SCGB2A1 OR MGB2 OR UGB3 OR LIPHC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99026127; PubMed=9806831;				
RA	Becker R.M., Darrow C., Zimonjic D.B., Popescu N.C., Watson M.A.,				
RA	Fleming T.P.;				
RT	"Identification of mammaglobin B, a novel member of the uteroglobin				
RT	gene family.";				
RL	Genomics 54:70-78(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99167354; PubMed=10066439;				
RA	Zhao C., Nguyen T., Yustifov T., Glasgow B.J., Lehrer R.I.;				
RA	"Lipophilin: human peptides homologous to rat prostatein.";				
RL	Biochem. Biophys. Res. Commun. 256:147-155(1999).				
RN	[3]				
RP	SEQUENCE OF 19-85.				
RC	TISSUE=Teats;				
RX	MEDLINE=98163342; PubMed=9504814;				
RA	Molloy M.P., Bolis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,				
RA	Willcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.;				
RT	"Establishment of the human reflex tear two-dimensional polycrylamide				
RT	gel electrophoresis reference map: new proteins of potential				
RT	diagnostic value.";				
RL	Electrophoresis 18:2811-2815(1997).				
RN	[4]				
RP	SEQUENCE OF 19-46 AND 60-78. AND MASS SPECTROMETRY.				
RX	MEDLINE=98385871; PubMed=9720917;				
RA	Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Qu X.-D., Martin D.,				
RA	Glasgow B.J.;				
RT	"Lipophilin, a novel heterodimeric protein of human tears.";				
RL	FEBS Lett. 432:163-167(1998).				
CC	- FUNCTION: MAY BIND ANDROGENS AND OTHER STEROIDS. MAY ALSO BIND				
CC	ESTRADIOL, A CHEMOTHERAPEUTIC AGENT USED FOR PROSTATE CANCER.				
CC	- SUBUNIT: HETERODIMER OF A LIPOPHILIN A AND A LIPOPHILIN C				
CC	(MAMMAGLOBIN B) MONOMER ASSOCIATED WITH HEAD TO HEAD.				
CC	- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, TRACHEA, KIDNEY, STERIOD				
CC	RESPONDING TISSUES (PROSTATE, TESTIS, UTERUS, BREAST AND OVARY),				
CC	AND SALIVARY GLAND.				
CC	- MASS SPECTROMETRY: MW=8854.94; METHOD=Electrospray; RANGE=19-95.				
CC	- SIMILARITY: BELONGS TO THE UTEROGLIBIN FAMILY. LIPOPHILIN				
CC	SUBFAMILY.				

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF071219; AAC79996.1; -.
DR EMBL: AJ224173; CA11865.1; -.
DR Genew: HGNC:7051; SCGB2A1.
DR MIM: 604398; -.
DR InterPro: IPR003627; Mamgb/prostatn.
DR InterPro: IPR000329; uteroglobin.
DR Pfam: PF01099; uteroglobin.1.
DR ProDom: PD029354; Mamgb/prostatn.1.
DR PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE: PS00404; UTEROGLOBIN_2; FALSE_NEG.
KW Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 95 MAMMAGLOBIN B.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 95 AA; 10884 MW; 0719738289F89FD CRC64;

Query Match
Best Local Similarity 100.0%; Score 496; DB 1; Length 95;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLLVLMALALLHCYAGSGCKLLEDVWEKTIINSISPEYKELQEFIDSDAAEAMG 60
DB 1 MKLLVLMALALLHCYAGSGCKLLEDVWEKTIINSISPEYKELQEFIDSDAAEAMG 60
OY 61 KFKQCFLNOSHRTLKNGFLMHTVYDSINCMKSN 95
DB 61 KFKQCFLNOSHRTLKNGFLMHTVYDSINCMKSN 95

RESULT 2
MGBA_HUMAN
ID MGBA_HUMAN STANDARD; PRT; 93 AA.
AC Q13296;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mammaglobin A precursor (Mammaglobin 1) (Secretoglobin family 2A
DE member 2).
OS SCGB2A2 OR MGB1 OR UGB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=breast;
RX MEDLINE=96223698; PubMed=8631025;
RA Watson M.A., Fleming T.P.;
RT "Mammaglobin, a mammary-specific member of the uteroglobin gene
RT family, is overexpressed in human breast cancer.";
RL Cancer Res. 56:860-865(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98147371; PubMed=9488047;
RA Watson M.A., Darlow C., Zimonjic D.B., Popescu N.C., Fleming T.;
RT "Structure and transcriptional regulation of the human mamaglobin
RT gene, a breast cancer associated member of the uteroglobin gene
RT family localized to chromosome 11q13.";
RL Oncogene 16:817-824(1998).
RN [3]
RP TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST
CC CANCER.
CC -1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC SUBFAMILY.
-----

```

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U33147; AAC50375.1; -.
DR EMBL: AF015224; AAC39608.1; -.
DR Genew: HGNC:7050; SCGB2A2.
DR MIM: 605562; -.
DR InterPro: IPR003627; Mamgb/prostatn.
DR InterPro: IPR000329; uteroglobin.
DR Pfam: PF01099; uteroglobin.1.
DR ProDom: PD029354; Mamgb/prostatn.1.
DR SMART: SM00096; UTRG; 1.
DR PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE: PS00404; UTEROGLOBIN_2; 1.
KW Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 93 POTENTIAL.
FT CARBOHYD 53 53 MAMMAGLOBIN A.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 93 AA; 10499 MW; 2896B8C43BF053B2 CRC64;

Query Match
Best Local Similarity 56.7%; Score 281; DB 1; Length 93;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

OY 1 MKLLVLMALALLHCYAGSGCKLLEDVWEKTIINSISPEYKELQEFIDSDAAEAMG 60
DB 1 MKLLVLMALALLHCYAGSGCKLLEDVWEKTIINSISPEYKELQEFIDSDAAEAMG 60
OY 61 KFKQCFLNOSHRTLKNGFLMHTVYDSINCM 92
DB 61 ELKECFLNQDETFLSNVEFMQIYDSICDL 92

RESULT 3
PSC3_RAT
ID PSC3_RAT STANDARD; PRT; 95 AA.
AC P02780; Q63463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Prostatic steroid-binding protein C3 chain precursor (Prostatein
DE peptide C3).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83082848; PubMed=6294095;
RA Parker M.G., White R., Hurst H., Needham M., Tilly R.;
RT "Prostatic steroid-binding protein. Isolation and characterization of
RT C3 gene.";
RL J. Biol. Chem. 258:12-15(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83238526; PubMed=6190812;
RA Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M.,
RA French F.S.;
RT "Isolation of two genomic sequences encoding the Mr = 14,000 subunit
RT of rat prostatein.";
RL J. Biol. Chem. 258:8861-8866(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165796; PubMed=1537831;
RA Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M.,
RA French F.S.;
RT "Response elements of the androgen-regulated C3 gene.";
-----

```

RL J. Biol. Chem. 267:4456-4466(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE-92218467; PubMed-1339454;
 RA Tan J.A., Marschke K.B., Ho K.C., Perry S.T., Wilson E.M.,
 RA French F.S.;
 RL J. Biol. Chem. 267:7958-7958(1992).
 RN [5]
 RP SEQUENCE OF 19-95.
 RX MEDLINE-81188769; PubMed-7014218;
 RA Peeters B., Rombauts W., Mous J., Heyns W.;
 RT Structural studies on rat prostatic binding protein. The primary
 RT structure of its glycosylated component C3";
 RL Eur. J. Biochem. 115:115-121(1981).
 CC -1- FUNCTION: PART OF PROSTATEIN WHICH IS THE MAJOR SECRETORY
 CC GLYCOPROTEIN OF VENTRAL PROSTATE GLAND. STEROID-BINDING PROTEIN;
 CC CAN BIND NON-POLAR STEROIDS, CHOLESTEROL AND A GROUP OF SMALL
 CC PROLINE-RICH PEPTIDES.
 CC -1- SUBUNIT: PROSTATEIN IS COMPOSED OF THREE DIFFERENT PEPTIDES CALLED
 CC C1, C2 AND C3. THESE FORM COVALENT C1:C3 (F) AND C2:C3 (S)
 CC HETERODIMERS WHOSE NONCOVALENT ASSOCIATION FORMS TETRAMERIC
 CC (C1:C3/C3:C2) PROSTATEIN MOLECULES.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: VENTRAL PROSTATE.
 CC -1- INDUCTION: ANDROGEN DEPENDENT, AS SHOWN BY THE DECREASE IN THE
 CC LEVEL OF THE PROTEIN FOLLOWING CASTRATION.
 CC -1- MISCELLANEOUS: C3 IS ENCODED BY TWO DIFFERENT GENES.
 CC -1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01263; CAA24577.1; -
 CC EMBL: M71245; AAA41965.1; -
 CC PIR: A03250; B0RT3.
 DR InterPro: IPR003627; Managb/prostatu.
 DR InterPro: IPR000329; Uteroglobin.
 DR Pfam: PF01099; Uteroglobin; 1.
 DR ProDom: PD029354; Managb/prostatu; 1.
 DR ProSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
 DR PROSITE: PS00404; UTEROGLOBIN_2; 1.
 KM Signal; Glycoprotein; Steroid-binding.
 FT SIGNAL 1 18
 FT CHAIN 19 95
 FT PROSTATIC STEROID-BINDING PROTEIN C3
 FT N-LINKED (GLCNAG. . .).
 FT CARBOHYD 35 35
 FT CONFLICT 53 53 D -> A (TN REF. 3).
 FT CONFLICT 79 79 G -> S (TN REF. 2).
 SQ SEQUENCE 95 AA; 10730 MW; F7F7F1A0C882E375 CRC64;
 Query Match 39.4%; Score 195.5; DB 1; Length 95;
 Best local Similarity 35.8%; Pred. No. 2,2e-13;
 Matches 34; Conservative 30; Mismatches 30; Indels 1; Gaps 1;
 QY 1 MKLILMILMALLLHCYAD-SGCKLEDVVEKTIINSDISIPEYKELLQETIDDAAEAM 59
 DB 1 MKLIVFLVLTIPICVAGSGCSILDEVIRGTINSTVTLHDYMKLVKPYVQHFEKAV 60
 QY 60 GKRCQCFNSHRTLNKFGMLMHTVYDSITWCNKS 94
 DB 61 KQFKQCFLDQTDKLTLENVGVMEAFINSESCQGPS 95
 RESULT 4
 ID YRS9_MYCTU STANDARD; PRT; 131 AA.
 AC 03301;

DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV2759C.
 GN RV2759C OR MT2829 OR MYV002.24C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham T., Gellies S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains." to the EMBL/Genbank/DBJ databases.
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0110 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: A008967; CA15555.1; -
 CC EMBL: AE007110; AKK47148.1; -
 CC TIGR: MT2829; -
 DR Tuberculin; RV2759C; -
 DR InterPro: IPR005342; PF03655.
 DR Pfam: PF03655; UPF0110; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 131 AA; 14372 MW; 9C12AE0AC674065B CRC64;
 Query Match 14.4%; Score 71.5; DB 1; Length 131;
 Best local Similarity 21.2%; Pred. No. 1.5;
 Matches 21; Conservative 18; Mismatches 35; Indels 25; Gaps 1;
 QY 5 MVLMLALLLHCYADSGCKLEDVVEKTIINSDISIPEYKELLQETIDDAAEAM 59
 DB 1 MIVDTSAIVAVISGESGAOVLEKALERSPNRSMSANVVELCAIMORDRPEISRLVRL 60
 QY 48 -----FRIDSDAAAEAMGKRCQCFNSHRTLNKFG 78
 DB 61 LDDYGIQVEAVDADQARVAQAQAYRDYGRSGHPARLNLG 99
 RESULT 5
 ID FTSY_RICPR STANDARD; PRT; 303 AA.
 AC 005948;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

FT DISULFID 28 28 INTERCHAIN (WITH C3) (PROBABLE).
 FT DISULFID 69 69 INTERCHAIN (WITH C3) (PROBABLE).
 FT DISULFID 92 92 INTERCHAIN (WITH C3) (PROBABLE).
 FT CONFLICT 26 26 MISSING (IN REF. 3).
 FT CONFLICT 88 88 I -> T (IN REF. 3).
 FT CONFLICT 96 96 VMLQINPGRFSEIN -> YGYK (IN REF. 3).
 SQ SEQUENCE 112 AA: 12828 MW: DA65A6A82B677864 CRC64;

Query Match 13.6%; Score 67.5; DB 1; Length 112;
 Best Local Similarity 27.0%; Pred. No. 32;
 Matches 27; Conservative 20; Mismatches 36; Indels 17; Gaps 5;

OY 4 LNWLMALALLHCYDPSG-----CKLEDM-VKRTINSISIPYKELQEFIDSDA 55
 DB 3 LSLCLTLIVVCCYANGQTHAGVQALQDVITITFLNPE--EELKRELEFDDAPPA 59
 OY 56 AEMAGKFKQCFNLQSHRTLNKFGIMHT--VYDSTWCNKK 93
 DB 60 VEANLKVRCI---NKIMYGRLSMGTSLVFTMLKCVK 95

RESULT 7

TRPC_ACICA STANDARD; PRT: 268 AA.
 AC P00911;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Indole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS).
 GN TRPC.
 OS Acinetobacter calcoaceticus.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 OC Acinetobacter.
 OX NCBI_TaxID=471;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88174326; PubMed=6599977;
 RA Kaplan J.B., Goncharoff P., Seibold A.M., Nichols B.P.;
 RT "Nucleotide sequence of the Acinetobacter calcoaceticus trpGDC gene cluster";
 RT Mol. Biol. Evol. 1:456-472(1984).
 CC -1- CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate - 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
 CC -1- PATHWAY: Tryptophan biosynthesis; fourth step.
 CC -1- SIMILARITY: BELONGS TO THE TRPC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M36636; AAA21905.1; -
 CC DR PIR: A01133; GMECC.
 CC DR HSSP: P00909; IPTI.
 CC DR InterPro: IPR001468; IGPS.
 CC DR Pfam: PF00218; IGPS.1.
 CC DR ProDom: PD001511; IGPS.1.
 CC DR PROSITE: PS00614; IGPS.1.
 CC KW Tryptophan biosynthesis; Lyase; Decarboxylase.
 SQ SEQUENCE 268 AA: 30216 MW: 8DBAC0505DA7527 CRC64;

Query Match 13.6%; Score 67.5; DB 1; Length 268;
 Best Local Similarity 26.4%; Pred. No. 8;
 Matches 19; Conservative 16; Mismatches 28; Indels 9; Gaps 2;

OY 12 LLHGYADSCLEEDVEKTIINSISIPYKELQEFIDSDAAEAMKFKQCFNLQSH 71
 DB 143 LTVACLSD---QGLEEKSTAFEDL-----DVLVEVHDEQLERLAKLSBQCLGYN 193
 OY 72 RLKNGFIMHT 83

DB 194 RNLKTFVDLNT 205

RESULT 8

RADC_HAEIN STANDARD; PRT: 234 AA.
 ID RADC_HAEIN
 AC P44952;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA repair protein radC homolog.
 GN RADC OR H10952.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Rd / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd";
 RT Science 269:496-512(1995).
 CC -1- FUNCTION: INVOLVED IN DNA REPAIR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RADC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U32776; AAC22613.1; -
 CC DR TIGR: H10952;
 CC DR InterPro: IPR001405; RADC.
 CC DR ProDom: PD007415; RADC.1.
 CC DR TIGRFAMS: TIGR00608; radC.1.
 CC DR PROSITE: PS01302; RADC.1.
 CC KW DNA repair; Complete proteome.
 SQ SEQUENCE 234 AA: 26774 MW: 25472EB289E5DA69 CRC64;

Query Match 13.4%; Score 66.5; DB 1; Length 234;
 Best Local Similarity 26.6%; Pred. No. 8;
 Matches 21; Conservative 15; Mismatches 26; Indels 17; Gaps 3;

OY 24 LLEDVWEKTIINSISIPYKELQEFIDSDAAEAMKFKQCFNLQSHRTLNKFGIMHT 83
 DB 103 LKODMLSTPIIND--PEYKRL--FLITELQHEEREVFWVLFDNQHRLIKRKERFLGT 156
 OY 84 VY-----DSIMCN 91
 DB 157 IYVSAYPREIKKALYCN 175

RESULT 9

IL15_MOUSE STANDARD; PRT: 162 AA.
 ID IL15_MOUSE
 AC P48346;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-DEC-1996 (Rel. 37, Last annotation update)

```

GN Interleukin-15 precursor (IL-15).
DE IL15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC/REJ X C57BL/6J; TISSUE=Bone marrow;
RX MEDLINE=95278940; PubMed=7759105;
RA Anderson D.M., Johnson L., Glaccum M.B., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., Valentine V., Kirsstein M.N., Shapiro D.N., Morris S.W.,
RA Grabstein K., Cosman D. ;
RT "Chromosomal assignment and genomic structure of IL15."
RL Genomics 25:701-706(1995).
CC -I- FUNCTION: CYTOKINE THAT STIMULATES THE PROLIFERATION OF T-
CC Lymphocytes. STIMULATION BY IL-15 REQUIRES INTERACTION OF IL-15
CC WITH COMPONENTS OF IL-2R, INCLUDING IL-2R BETA AND PROBABLY IL-2R
CC GAMMA BUT NOT IL-2R ALPHA.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE IL-15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14332; AAF5377.1; -.
CC MCD, MGI:103014; IL15.
DR InterPro: IPR003443; Interleukin_15.
DR Pfam: PF02372; IL15; 1.
RW Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT PROPEP 30 48 POTENTIAL.
FT CHAIN 49 162 INTERLEUKIN-15.
FT DISULFID 83 133 POTENTIAL.
FT DISULFID 90 136 POTENTIAL.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 162 AA; 18593 MW; 66C971496CEBF296 CAC64;

Query Match 13.2%; Score 65.5; DB 1; Length 162;
Best Local Similarity 34.0%; Pred. No. 7.6;
Matches 18; Conservative 6; Mismatches 10; Indels 19; Gaps 2;

QY 18 ADSGKLLDEWVKRTINDISTPEYKELLOERISDAAAEANGKRGKCFINS 70
DB 129 AEGSGKECELEERT-----PTEFLQSEI-----RIVQMFINTS 162
ID VE1_HPV08 STANDARD: PRT; 603 AA.
AC P06420;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Replication protein E1.
GN E1.
OS Human papillomavirus type 8.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
CX NCBI_TaxID=10579;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86200410; PubMed=3009874;
RA Fuchs P.G., Iftner T., Weninger J., Pfister H.;
RT "Epidermodysplasia verruciformis-associated human papillomavirus 8:
RT genomic sequence and comparative analysis."

```

```

RL J. Virol 58:626-634(1986).
CC -I- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M12737; -; NOT_ANNOTATED_CDS.
CC DR PIR; A03656; W1ML8.
CC DR InterPro; IPR001177; Pap11low_E1.
CC DR Pfam; PF00519; E1; 1.
CC DR Pfam; PF00524; E1_N_1.
CC DR Early protein; DNA replication; 'helicase; ATP-binding; DNA-binding;
CC Nuclea protein.
CC KW NP_BIND 431 438 ATP (POTENTIAL).
CC FT SEQUENCE 603 AA; 68821 MW; 0813660098DA68AD CRC64;
CC -----
CC Query Match 13.2%; Score 65.5; DB 1; Length 603;
CC Best Local Similarity 22.2%; Pred. NO. 30;
CC Matches 22; Conservative 21; Mismatches 23; Indels 33; Gaps 4;
CC -----
Oy 19 DSGCGLLEDMYEKRTNSDISIP-----EYKELLOERLSDPAAMEAGK 61
Db 112 DSGVELTLNNEADVSHVEVPAIDRSPEDEGSCALDIDYTLALR---SSWTKATLMAK 168
Oy 62 FKOCFLN-----QSHRTLKNFGLM---MHTVYDS 87
Oy 169 FKEARGDGFNELTRQFKSYKTCNNWVVAAYAVHDVYES 207
Db -----
RESULT 11
ID GLT1_YEAST STANDARD; PRT; 2144 AA.
AC 012680; 012290;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutamate synthase [NADPH] precursor (EC 1.4.1.13) (NADPH-GOGAT).
DS GLT1 OR YDL171C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RX [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CN36;
RX MEDLINE=97082505; PubMed=8923741;
RA Filletici P., Martegani M.P., Valenzuela L., Gonzalez A., Ballario P.;
RT "Sequence of the GLT1 gene from Saccharomyces cerevisiae reveals the
RT domain structure of yeast glutamate synthase.";
RL Yeast 12:1359-1366(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Pohl T.M.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: 2 L-glutamate + NADP(+) = L-glutamine + 2-
CC oxoglutarate + NADPH.
CC -I- COFACTOR: BINDS A 3FE-4S CLUSTER; FAD AND FMN.
CC -I- SUBUNIT: HOMOTRIMER.
CC -I- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



```

ID DYNL_ARATH STANDARD: PRT: 611 AA.
AC P42697;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dynamin-like protein.
OS Arabidopsis thaliana (Mouse-ear cress); Embryophyta: Tracheophyta;
OC Eukaryota: Viridiplantae: Streptophyta;
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
OC eusids II: Brassicales; Brassicaceae: Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=leaf;
RC Huang J., Goodman H.M., Yoon H.W., Yoon J.-H., Cho M.J.;
RT Cloning and characterization of a cDNA clone encoding dynamin-like
RT GTP binding protein in Arabidopsis thaliana.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L36939; AAB63528.1; -
DR InterPro: IPR001401; Dynamin.
DR InterPro: IPR000375; Dynamin_central.
DR InterPro: IPR003130; GED.
DR Pfam: PF00350; dynamin_1.
DR Pfam: PF01031; dynamin_2; 1.
DR Pfam: PF02212; GED; 1.
DR PRIN2: PR00195; DYNAMIN.
DR SMART: SM0053; DYNC; 1.
DR SMART: SM00302; GED; 1.
DR PROSITE: PS00410; DYNAMIN; 1.
KW Motor protein; GTP-binding.
FT NP_BIND 41 48 GTP (POTENTIAL).
FT NP_BIND 142 146 GTP (POTENTIAL).
FT NP_BIND 211 214 GTP (POTENTIAL).
SQ SEQUENCE 611 AA; 68509 MW; EDE98B/9B1BC03D5 CRC64;

Query Match 13.0%; Score 64.5; DB 1; Length 611;
Best local Similarity 20.7%; Pred. No. 38;
Matches 17; Conservative 21; Mismatches 23; Indels 21; Gaps 3;

QY 24 LLEDVKEKINSISPEYKELLOEPIIDSDAAEAMCKFK-----QC----- 65
DB 432 ILKDLVHKSVEYVELKQYPALREY--TNAALIESLDKMRGSKKATLQLVDMCSYLTV 489
QY 66 -FLNOSHRTLKNFGMLMHTYD 86
DB 490 DFRRLKLPQDVEKGNPHTSIFD 511

RESULT 15
ID DOT1_YEAST STANDARD: PRT: 582 AA.
AC 004089;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Disrupter of telomere silencing protein 1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes;
OC Saccharomycetales: Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.

```

```

RX MEDLINE-98429517; PubMed-9755194;
RA Singer M.S., Kahana A., Wolf A.J., Weisinger L.L., Peterson S.E.,
RA Goggin C., Mahowald M., Gottschling D.E.;
RT "Identification of high-copy disruptors of telomeric silencing in
RT Saccharomyces cerevisiae.";
RL Genetics 150:613-632(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Dierlich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Hunkele-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Botstein D., Davis R.W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Has a role in telomere silencing, which is the
CC repression of chromatin structure which leads to a stop in the
CC transcription of nearby genes.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33007; AAB64868.1; -
DR SGD: S0002848; DOT1.
KW Transcription regulation.
SQ SEQUENCE 582 AA; 66201 MW; 05CAA6A8F8CBAB9A CRC64;

Query Match 12.8%; Score 63.5; DB 1; Length 582;
Best local Similarity 25.7%; Pred. No. 46;
Matches 19; Conservative 13; Mismatches 15; Indels 27; Gaps 3;

QY 10 AAILLHCYADSGCKLLEDVKEKINSISPEYK-----LLEPIID 51
DB 409 AALECCALSGCEIMDA-----SDLTILQYEELKRRCKLYGMLNVEFSLKRSFVD 462
QY 52 SDAAEAMCKFKOC 65
DB 463 NNRYAEILI---PQC 473

Search completed: January 25, 2003, 13:31:47
Job time : 85 secs

```


09GK61
AC 09GK61 PRELIMINARY; PRT: 93 AA.
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Lipophilin CS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SUBMAXILARY;
RA Zhao C., Nguyen T.X., Lehrer R.I.;
RT "Rabbit Lipophilin".
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF308620; AAG42808.1; -
DR InterPro: IPR003627; Mamgb/prostatn.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR PRINTS: PR00486; UTEROGLBLN.
DR ProDom: PD029354; Mamgb/prostatn; 1.
DR SMART: SM0096; UTG; 1.
SQ SEQUENCE 93 AA; 10609 MW; E2B015AD31B2249 CRC64;

Query Match 46.8%; Score 232; DB 6; Length 93;
Best Local Similarity 50.0%; Pred. No. 21e-17;
Matches 45; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

OY 1 MKLMLVLMALLHCAVDSCKLLEDMVEKTIINSISPEYKELLOEFTIDSDAAAEAWG 60
DB 1 MKVIVLMLALPLCYAGSGCPVEKMKVKTINSNVSTAEYIDVKNYINDETELAVV 60
OY 61 KFKOCFLNOSHRTLKNFGLMHTVYDSIWC 90
DB 61 EFKNCFLOSSEETLRNVEMETIYNSKLC 90

RESULT 3

09GK64
ID 09GK64 PRELIMINARY; PRT: 93 AA.
AC 09GK64;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Lipophilin CL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LACRIMAL GLAND;
RA Zhao C., Nguyen T.X., Lehrer R.I.;
RT "Rabbit Lipophilin".
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF308617; AAG42805.1; -
DR InterPro: IPR003627; Mamgb/prostatn.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR ProDom: PD029354; Mamgb/prostatn; 1.
SQ SEQUENCE 93 AA; 10350 MW; 9F9206C4372804D CRC64;

Query Match 46.2%; Score 229; DB 6; Length 93;
Best Local Similarity 45.6%; Pred. No. 4.4e-17;
Matches 41; Conservative 26; Mismatches 23; Indels 0; Gaps 0;

OY 1 MKLMLVLMALLHCAVDSCKLLEDMVEKTIINSISPEYKELLOEFTIDSDAAAEAWG 60
DB 1 MKVIVLMLALPLCYAGSGCVILLESVDKTIIDPSVSEVDITYTLOKYLITDAKVALE 60
OY 61 KFKOCFLNOSHRTLKNFGLMHTVYDSIWC 90

DB 61 ELKOCFLSOSNETLANVKVLEAVFDSLXC 90

RESULT 4

09GK62
ID 09GK62 PRELIMINARY; PRT: 93 AA.
AC 09GK62;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Lipophilin CP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RA Zhao C., Nguyen T.X., Lehrer R.I.;
RT "Rabbit Lipophilin".
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF308619; AAG42807.1; -
DR InterPro: IPR003627; Mamgb/prostatn.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR ProDom: PD029354; Mamgb/prostatn; 1.
SQ SEQUENCE 93 AA; 10332 MW; C9DC35B17D372F32 CRC64;

Query Match 45.8%; Score 227; DB 6; Length 93;
Best Local Similarity 50.0%; Pred. No. 7.2e-17;
Matches 45; Conservative 16; Mismatches 29; Indels 0; Gaps 0;

OY 1 MKLMLVLMALLHCAVDSCKLLEDMVEKTIINSISPEYKELLOEFTIDSDAAAEAWG 60
DB 1 MKCVIALMLALPLCYAGSGCOLLDDMYTKTLDQSITLDYINFFKNLSSGAAEKAVK 60
OY 61 KFKOCFLNOSHRTLKNFGLMHTVYDSIWC 90
DB 61 DFKOCFLMSNETLNNIKVLETVYNSPFC 90

RESULT 5

09JHB9
ID 09JHB9 PRELIMINARY; PRT: 95 AA.
AC 09JHB9;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Prostatic steroid binding protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=84057754; PubMed=6685625;
RX Hurst H.C., Parker M.G.;
RT "Rat prostatic steroid binding protein: DNA sequence and transcript maps of the two C3 genes.";
RL EMBO J. 2:769-774(1983).
DR EMBL: V01260; CAB75892.1; -
DR EMBL: V01261; CAB75892.1; JOINED.
DR EMBL: V01262; CAB75892.1; JOINED.
DR InterPro: IPR003627; Mamgb/prostatn.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR ProDom: PD029354; Mamgb/prostatn; 1.
SQ SEQUENCE 95 AA; 10622 MW; B209F1BE177C52A4 CRC64;

Query Match 38.6%; Score 191.5; DB 11; Length 95;
Best Local Similarity 34.7%; Pred. No. 4.7e-13;
Matches 33; Conservative 31; Mismatches 30; Indels 1; Gaps 1;


```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0029 FAMILY.
DR EMBL; Z88978; CAB1676.1;
DR InterPro; IPR001498; UPF0029.
DR Pfam; PF01205; UPF0029; 1.
KM Hypothetical protein.
FT DOMAIN 276 280 POLY-LYS.
SQ SEQUENCE 280 AA; 31485 MW; 90PDE06A14091219 CRC64;

Query Match 15.4%; Score 76.5; DB 3; Length 280;
Best Local Similarity 30.3%; Pred. No. 3.2;
Matches 23; Conservative 10; Mismatches 28; Indels 15; Gaps 3;

OY 24 LLEDNWEKTIINSISPEYKELLOEFIDSD---AAAEAMCKFRQ-----CFINSHRT 73
DB 68 LADVDLVSATGVCVFSTYMDLKLKELVDIDAEGAAERESKLOESDKETPPVMLNRSHY 127
OY 74 LKNFGMLMHTVYDSIW 89
DB 128 AKT-----PEIODEPW 138

RESULT 10
OY 09A622 PRELIMINARY; PRT; 212 AA.
AC 09A622;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE D-alanyl-D-alanine dipeptidase.
GN CC2273.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=158892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RA MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Pladke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005897; AKK24244.1;
DR TIGR; CC2273;
DR InterPro; IPR000755; Peptidase_M45.
DR Pfam; PF01427; Peptidase_M45; 1.
DR ProDom; PD013478; Peptidase_M45; 1.
KM Complete proteome.
SQ SEQUENCE 212 AA; 23874 MW; 82846D0F44F07028 CRC64;

Query Match 15.3%; Score 76; DB 16; Length 212;
Best Local Similarity 29.9%; Pred. No. 2.7;
Matches 26; Conservative 13; Mismatches 16; Indels 32; Gaps 5;

OY 18 ADGCGKLEMEVKT-----INSDISPEYKE---LLOEFIDSDAAAEAMG 60
DB 19 AETGDKLASLVVYTTTTCIKIDIRKACANNMGIPLYEKSAVYLR-----PAEALG 73
OY 61 KFKOCEFLNOSHRTL--KNFGMLMHTVY 85

```

```

DB 74 RI-----HRLAAKGYGLITHDAY 92

RESULT 11
OY 09CAN6 PRELIMINARY; PRT; 590 AA.
AC 09CAN6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 66.1 kDa protein.
GN F16M19.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.U., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Malt R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Paj G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RT Nature 408:816-820(2000).
DR EMBL; AC010795; AAG51617.1;
DR InterPro; IPR002885; PPR.
DR Pfam; PF01535; PPR; 15.
DR TIGRPFam; TIGR00756; PPR; 13.
KM Hypothetical protein.
SQ SEQUENCE 590 AA; 66071 MW; 596067345518244F CRC64;

Query Match 14.8%; Score 73.5; DB 10; Length 590;
Best Local Similarity 32.3%; Pred. No. 16;
Matches 21; Conservative 8; Mismatches 23; Indels 13; Gaps 3;

OY 13 LHCYADSG-----CKLEMEVKTINSISPEYKELLOEFIDSDAAAF-----MCK 61
DB 291 LILCCLNYGRWSDASRLISMLEKINPDLVF--FNALIDAFVKEGKLYAEKLYDEMKV 348
OY 62 FKOCF 66
DB 349 SKHCF 353

RESULT 12
OY 057483 PRELIMINARY; PRT; 1688 AA.
AC 057483;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Voltage-dependent L-type calcium channel, ALPHA-1S subunit
DE (FGALPHA1S).
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.

```

OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKETALET MUSCLE;
RX MEDLINE=98411359; PubMed=9738021;
RA Zhou J., Cribbs L., Yi J., Shirokov R., Perez-Reyes E., Rios E.;
RT "Molecular cloning and functional expression of a skeletal muscle
dihydropyridine receptor from Rana catesbeiana.";
RL J. Biol. Chem. 273:25503-25509(1998).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIANE THE
ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CONTRACTION, GENE EXPRESSION, CELL MOTILITY, CELL DIVISION AND
CELL DEATH. THE ISOFORM ALPHA-1S GIVES RISE TO L-TYPE CALCIUM
CURRENTS.
CC -1- FUNCTION: LONG-LASTING (L-TYPE) CALCIUM CHANNELS BELONG TO THE
"HIGH-VOLTAGE ACTIVATED (HVA)" GROUP. THEY ARE BLOCKED BY
DIHYDROPYRIDINES (DHP), PHENYLAALKYLAMINES, BENZOTHAZEPINES, AND B
OMEGA-AGATOXIN-IIA (OMEGA-AGA-IIA). THEY ARE HOWEVER INSENSITIVE
TO OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA
(OMEGA-AGA-IVA). CALCIUM CHANNELS CONTAINING THE ALPHA-1S SUBUNIT
PLAY AN IMPORTANT ROLE IN EXCITATION-CONTRACTION COUPLING IN
SKELETAL MUSCLE (BY SIMILARITY).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY. AN
ADDITIONAL GAMMA SUBUNIT IS PRESENT ONLY IN SKELETAL MUSCLE L-TYPE
CHANNEL (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE SPECIFIC.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- DOMAIN: THE LOOP BETWEEN REPEATS II AND III INTERACTS WITH THE
RYANODINE RECEPTOR, AND IS THEREFORE IMPORTANT FOR CALCIUM RELEASE
FROM THE ENDOPLASMIC RETICULUM NECESSARY FOR MUSCLE CONTRACTION
(BY SIMILARITY).
CC -1- PTM: PHOSPHORYLATION BY CA2P STIMULATES THE CALCIUM CHANNEL
FUNCTION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
FAMILY.
CC EMBL: AF037625; AAC36126.1; -;
DR InterPro: IPR001682; Ca_Na_pore.
DR InterPro: IPR002111; Ca_channel_TrpL.
DR InterPro: IPR002077; Ca_channel.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR000636; M+channel_nlg.
DR Pfam: PF00520; Ion_trans_4.
DR PRINTS: PR00167; CACHANNEL.
DR PRINTS: PR01630; LVDCALPHAL.
DR PRINTS: PR01634; LVDCALPHALS.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA. UNKNOWN_1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation.
FT REPEAT 38 335 I.
FT REPEAT 417 663 II.
FT REPEAT 784 1066 III.
FT REPEAT 1103 1371 IV.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 52 70 S1 OF REPEAT I (POTENTIAL).
FT DOMAIN 71 88 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 89 108 S2 OF REPEAT I (POTENTIAL).
FT DOMAIN 109 120 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 121 139 S3 OF REPEAT I (POTENTIAL).
FT DOMAIN 140 158 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 159 177 S4 OF REPEAT I (POTENTIAL).

FT DOMAIN 178 196 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 197 216 S5 OF REPEAT I (POTENTIAL).
FT DOMAIN 217 307 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 308 332 S6 OF REPEAT I (POTENTIAL).
FT DOMAIN 333 431 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 432 450 S1 OF REPEAT II (POTENTIAL).
FT DOMAIN 451 465 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 466 485 S2 OF REPEAT II (POTENTIAL).
FT DOMAIN 486 493 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 494 512 S3 OF REPEAT II (POTENTIAL).
FT DOMAIN 513 522 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 523 541 S4 OF REPEAT II (POTENTIAL).
FT DOMAIN 542 560 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 561 580 S5 OF REPEAT II (POTENTIAL).
FT DOMAIN 581 635 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 636 660 S6 OF REPEAT II (POTENTIAL).
FT DOMAIN 661 797 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 798 816 S1 OF REPEAT III (POTENTIAL).
FT DOMAIN 817 832 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 833 852 S2 OF REPEAT III (POTENTIAL).
FT DOMAIN 853 864 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 865 883 S3 OF REPEAT III (POTENTIAL).
FT DOMAIN 884 890 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 891 909 S4 OF REPEAT III (POTENTIAL).
FT DOMAIN 910 928 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 929 948 S5 OF REPEAT III (POTENTIAL).
FT DOMAIN 949 1038 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1039 1063 S6 OF REPEAT III (POTENTIAL).
FT DOMAIN 1064 1116 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1117 1135 S1 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1136 1150 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1151 1170 S2 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1171 1178 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1179 1197 S3 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1198 1218 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1219 1237 S4 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1238 1256 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1257 1275 S5 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1277 1343 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1344 1368 S6 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1369 1688 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 22 28 POLY-PRO.
FT DOMAIN 438 441 POLY-ILE.
FT TRANSMEM 561 567 POLY-LEU.
FT DOMAIN 744 749 POLY-GLU.
FT TRANSMEM 1552 1559 POLY-GLU.
FT DOMAIN 355 372 BINDING TO THE BETA SUBUNIT (BY
SIMILARITY).
FT SITE 290 290 CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
FT SITE 612 612 CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
FT SITE 1012 1012 CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
FT SITE 1310 1310 CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
FT BINDING 986 1075 TO DIHYDROPYRIDINES (BY SIMILARITY).
FT BINDING 1324 1390 TO DIHYDROPYRIDINES (BY SIMILARITY).
FT BINDING 1336 1379 TO PHENYLAALKYLAMINES (BY SIMILARITY).
FT MOD_RES 1379 1379 PHOSPHORYLATION (BY CA2P) (POTENTIAL).
FT CA_BIND 1397 1408 BY SIMILARITY.
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1688 AA; 192421 MW; 58DA06B38823017A CRC64;
Query Match 14.7%; Score 73; DB 13; Length 1688;
Best Local Similarity 24.1%; Pred. No. 57;
Matches 28; Conservative 16; Mismatches 36; Indels 36; Gaps 5;
OY 3 LNWYMLAALLLHCYADSGCKLEDEWEKTI--NSDISPEYKELLOEFIDSD--AAAE 57
DB 194 LNLIALLVLFILITIAVGELEFSGKMKHTCYCFKDTIDTAT-----VDNEKPAPCS 244

QY 58 AMCKEFCPLNOSH-----RTLKNFCMLMHTVYDSI-----WCN 91
: : : : :
DB 245 STGCGGCSINSSECRGMPGPNNGITHEFDNFGFAMLTYYQCITMEGMEVLVWVN 300

RESULT 13

Q9LV86 PRELIMINARY: PRT: 487 AA.
AC Q9LV86: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
OS Genomic DNA, chromosome 5, p1 clone:MK3.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eurosid II: Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_TaxId=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA:
RX MEDLINE=20181125; PubMed=10718197;
RA Tabata S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Kotani H.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty p1 and YAC
RT clones.";
RL DNA Res. 7:31-63(2000).
DR EMBL: AB019236; BAA97303.1;
SQ SEQUENCE 487 AA: 54862 MW: 632BA9E298807D27 CRC64;

Query Match 14.4% Score 71.5; DB 10; Length 487;
Best Local Similarity 22.7%; Pred. No. 21;
Matches 17; Conservative 16; Mismatches 33; Indels 9; Gaps 1;
QY 14 LHCYASCCLEDMVEKTNISDISIPEYKELLOEFD-----DAAAEAMGKFKQ 64
: : : : :
DB 366 VEECYMVSVIEEFTDSLNEVLLPEQYAFEEFYKESDKAMAYDTAOALEKAKE 425
: : : : :
QY 65 CFLNOSRRLKNEGL 79
: : : : :
DB 426 GLSEETKKAQEMRL 440
: : : : :

RESULT 14

Q9ZGB8 PRELIMINARY: PRT: 303 AA.
AC Q9ZGB8: 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
OS Cell division protein ftsy.
GN FTSY OR RC1205.
OS Rickettsia conorii.
OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales:
OC Rickettsiaceae: Rickettsiidae: Rickettsia.
OX NCBI_TaxId=781;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-MALISH 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR EMBL: AE008667; AAL03743.1;
DR InterPro: IPR004390; FtsY.
DR InterPro: IPR000897; SRP54.
DR InterPro: IPR004781; SRP_dock.
DR Pfam: PF00448; SRP54; 1.
DR Pfam: PF02881; SRP54_N; 1.
DR ProDom: PD000819; SRP54; 1.

DR TIGRFAMS: TIGR00961; 3a0501s03; 1.
DR TIGRFAMS: TIGR00064; ftsy; 1.
DR PROSITE: PS00300; SRP54; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 303 AA: 33137 MW: B76FCBCFAC36CGB CRC64;

Query Match 14.3% Score 71; DB 16; Length 303;
Best Local Similarity 33.3%; Pred. No. 14;
Matches 20; Conservative 13; Mismatches 17; Indels 10; Gaps 2;

QY 19 DSGCKLLEDVVEKTNISDISIPEYKELLOE-----IDSNAAAEAMGKFKPLNOS 70
: : : : :
DB 33 DAG-TLEELELLISSDMSVLVVTNIEEKKVKEDKDEIDSTVKEALAKLIEQLSKS 90
: : : : :

RESULT 15

Q8TOR1 PRELIMINARY: PRT: 767 AA.
ID Q8TOR1
AC Q8TOR1: 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN CH12888p.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephydroidea: Drosophilidae: Drosophila.
OX NCBI_TaxId=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclad J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY069114; AAL39259.1;
SQ SEQUENCE 767 AA: 86535 MW: D60DEF6719F065E9 CRC64;

Query Match 14.3% Score 71; DB 5; Length 767;
Best Local Similarity 28.1%; Pred. No. 39;
Matches 25; Conservative 14; Mismatches 24; Indels 26; Gaps 4;

QY 19 DSGCKLLEDVVEKTNISDISIPEYK-----LLOEFTDSNAAAEAMGKFK 63
: : : : :
DB 25 ESDIEHLEORLEKTIKLCNVAVDSGKEYVKNQSAFAMSLMDLQOHFLDNKNAHNAUGKLI 84
: : : : :
QY 64 QCF-----LNOSHRP-LKNGGLM 81
: : : : :
DB 85 HCFQEMKFFHTILLDQASRTVLKLNLSVFF 113
: : : : :

Search completed: January 25, 2003, 13:34:11
Job time : 133 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:18:14 ; Search time 139 Seconds
(without alignments)
91.071 Million cell updates/sec

Title: US-09-806-302a-2
Perfect score: 496
Sequence: 1 MKLWLMIALLLHCYADS.....NFGIMHTYDSIMCKSN 95

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_101002: *
1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT: *
2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT: *
3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT: *
4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT: *
5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT: *
6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT: *
7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT: *
8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT: *
9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT: *
10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT: *
11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT: *
12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT: *
13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT: *
14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT: *
15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT: *
16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT: *
17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT: *
18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT: *
19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT: *
20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT: *
21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT: *
22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT: *
23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496	100.0	95	AAW35804	Human endometrial
2	496	100.0	95	AAU02590	A human mammary
3	496	100.0	95	AAU03769	Human endometrial
4	496	100.0	95	AAU92226	Human endometrial
5	496	100.0	95	AAU92237	Mammary gland
6	496	100.0	95	AAU65394	Human 5' EST relat
7	496	100.0	95	AAU31682	An endometrial spe
8	496	100.0	95	AAU02055	Protein of human l
9	496	100.0	95	AAU09635	Human endometrial
10	496	100.0	108	AAU60038	Human endometrium

11	390	78.6	74	19	AAU61649	Non-ocular disease
12	283	57.1	93	23	AAU83620	Human PRO protein
13	281	56.7	93	18	AAU10179	Mammary-specific s
14	281	56.7	93	19	AAU59777	Amino acid sequenc
15	281	56.7	93	19	AAU48432	Mammary protein
16	281	56.7	93	20	AAU01718	Mammary protein, a mam
17	281	56.7	93	21	AAU13786	Human mammary
18	281	56.7	93	21	AAU48422	Amino acid sequenc
19	281	56.7	93	22	AAU07517	Human mammary
20	281	56.7	93	22	AAU51127	Human mammary
21	281	56.7	410	22	AAU33359	Human breast cancer
22	281	56.7	743	22	AAU33358	Human breast cancer
23	281	56.7	1095	22	AAU33357	Human breast cancer
24	280	56.5	93	22	AAU07531	Human mammary
25	279	56.2	93	22	AAU07529	Human mammary
26	277	55.8	93	22	AAU07534	Human mammary
27	277	55.8	93	22	AAU07535	Human mammary
28	277	55.8	93	22	AAU07536	Human mammary
29	273	55.0	93	22	AAU07530	Human mammary
30	272	54.8	93	22	AAU07532	Human mammary
31	269	54.2	93	22	AAU07533	Human mammary
32	265.5	53.5	90	22	AAU07528	Human mammary
33	265.5	53.5	90	22	AAU07527	Human mammary
34	206	41.5	74	21	AAU48424	Amino acid sequenc
35	200	40.3	220	22	AAU22141	Ra12-mammary
36	95	19.2	33	19	AAU48433	Mammary protein
37	77	15.5	21	22	AAU51114	Human mammary
38	77	15.5	21	22	AAU51124	Mammary protein
39	77	15.5	21	22	AAU51126	Mammary protein
40	77	15.5	30	19	AAU48435	Mammary protein
41	72	14.5	20	22	AAU51132	Human mammary
42	71	14.3	828	22	AAU62649	Drosophila melanog
43	71	14.3	1187	22	AAU63669	Drosophila melanog
44	68.5	13.8	619	22	AAU97025	Human colon carcin
45	68	13.7	55	22	AAU51128	Human mammary

ALIGNMENTS

RESULT 1	
AAW35804	AAW35804 standard; Protein; 95 AA.
ID	AAW35804
AC	AAW35804:
DT	27-MAR-1998 (first entry)
XX	Human endometrial specific steroid-binding factor III.
DE	
XX	
KW	Endometrial specific steroid-binding factor III; ESF III; human;
KW	Clara cell secretory protein; endometrium;
KW	phospholipase A2 inhibitor; polychlorinated biphenyl; antiagregant;
KW	inflammation; asthma; rhinitis; cystic fibrosis; airway disease;
KW	neoplasia; atopy; therapy; diagnosis.
OS	Homo sapiens.
XX	
FH	Key
FT	Peptide
FT	1..21
FT	/label= Sig_Peptide
FT	Protein
FT	22..95
FT	/label= Mat_Protein
XX	
XX	W09734997-A1.
XX	
XX	25-SEP-1997.
XX	
PF	21-MAR-1996; 96MO-US03857.
XX	
XX	21-MAR-1996; 96MO-US03857.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.

XX Gentz RL, Ni J, Yu G;
PI N-PSDB; AAX36138.
XX WPI: 1997-480206/44.
DR N-PSDB; AAT94832.
XX Human endometrial specific steroid-binding factor I, II and III -
PT used to treat inflammation, asthma, rhinitis, cystic fibrosis,
XX airway disease, neoplasia, atopy etc.
PS Claim 19; Page 65; 92pp; English.
XX This sequence comprises human endometrial specific steroid binding
XX factor III (ESF III), a protein that inhibits phospholipase A2
XX activity, binds to polychlorinated biphenyl compounds, reduces
XX foreign protein antigenicity, inhibits monocyte and neutrophil
XX chemotaxis and phagocytosis, inhibits platelet aggregation,
XX regulates eicosanoid levels in the human uterus and controls the
XX growth of endometrial cells. The amino acid sequence was deduced
XX from a cDNA clone (see AAT94832) derived from a human endometrial
XX tumour. ESF I (see AAW35802) and ESF II (see AAW35803) are also
XX claimed. Human ESF III has about 36% identity with rat prostatic
XX steroid-binding protein. Recombinant ESF I, II and III can be
XX expressed in host cells for use in claimed methods (a) for treating
XX a patient in need of ESF I, II or III (including expression of the
XX polypeptide in vivo) and (b) for identifying compounds which bind
XX to and inhibit activation of the ESF polypeptide. hESF I, II and
XX CC III may be used to treat inflammation, asthma, rhinitis, cystic
XX CC fibrosis, airway disease, neoplasia and atopy.
SQ Sequence 95 AA;
Query Match 100.0%; Score 496; DB 18; Length 95;
Best Local Similarity 100.0%; Pred. No. 4,5e-48;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MKLLVLMALALLHCYADSGCKLLEDVMEKTNDSISPEYKELLOEFIDSDAAAEAMG 60
DB 1 MKLLVLMALALLHCYADSGCKLLEDVMEKTNDSISPEYKELLOEFIDSDAAAEAMG 60
OY 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIWCNMSN 95
DB 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIWCNMSN 95
RESULT 2
AAY02590 standard; Protein: 95 AA.
XX ID AAY02590 standard; Protein: 95 AA.
XX AC AAY02590;
XX DT 26-JUL-1999 (first entry)
XX DE A human mammary globin homologue (HMH).
XX KW Human mammary globin homologue; HMH: antagonist; neoplastic disorder;
XX KM adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
XX KW teratocarcinoma; endometrios.
XX OS Homo sapiens.
XX PN MO9919487-A1.
XX PD 22-APR-1999.
XX PF 14-OCT-1998; 98MO-US21729.
XX PR 16-OCT-1997; 97US-0951750.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Hillman JL, Murry LE, Shah P;
XX PT

DR WPI: 1999-302531/25.
XX N-PSDB; AAX36138.
XX New human mammary globin homologue (HMH), useful for diagnosing, treating
PT or preventing disorders associated with expression of HMH
XX Claim 1; Fig 1A-B; 63pp; English.
XX The present sequence represents a human mammary globin homologue (HMH).
XX Antagonists of the HMH polypeptide can be used to treat neoplastic
XX disorders including adenocarcinoma, leukemia, lymphoma, melanoma,
XX myeloma, sarcoma and teratocarcinoma. A vector expressing the
XX complement of the polynucleotide encoding HMH may be administered
XX to a subject to treat or prevent neoplastic disorders or endometrios.
XX Antipodas which bind HMH may also be used in the diagnosis of
XX conditions or diseases characterized by expression of HMH, or in assays
XX to monitor patients being treated with HMH, agonists, antagonists or
XX inhibitors. Polynucleotides encoding HMH may also be used diagnostically
XX to detect and quantitate gene expression in biopsied tissues. With
XX respect to cancer a relatively high amount of transcript may indicate a
XX predisposition for the development of disease. The nucleic acid sequences
XX CC which encode HMH may also be used to generate hybridization probes useful
XX CC for mapping the naturally occurring genomic sequence. HMH, and its
XX CC fragments/variants can be used for screening libraries of compounds in
XX drug screening techniques.
SQ Sequence 95 AA;
Query Match 100.0%; Score 496; DB 20; Length 95;
Best Local Similarity 100.0%; Pred. No. 4,5e-48;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MKLLVLMALALLHCYADSGCKLLEDVMEKTNDSISPEYKELLOEFIDSDAAAEAMG 60
DB 1 MKLLVLMALALLHCYADSGCKLLEDVMEKTNDSISPEYKELLOEFIDSDAAAEAMG 60
OY 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIWCNMSN 95
DB 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIWCNMSN 95
RESULT 3
AAB03769 standard; Protein: 95 AA.
XX ID AAB03769 standard; Protein: 95 AA.
XX AC AAB03769;
XX DT 06-OCT-2000 (first entry)
XX DE Human endometrial specific steroid-binding factor III protein sequence.
XX KW Endometrial specific steroid-binding factor; human; hESF; inflammation;
XX KM asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; atopy;
XX KW eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.
XX OS Homo sapiens.
XX PN US6066724-A.
XX PD 23-MAY-2000.
XX PF 21-MAR-1997; 97US-0821451.
XX PR 21-MAR-1996; 96US-0014724.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Yu G, Gentz R, Ni J;
XX WPI: 2000-375600/32.
XX N-PSDB; AAA59730.
XX Novel gene encoding human endometrial specific steroid-binding factor
PT

CC cancer, immune disorders, autoimmune disease, acquired immune deficiency
 CC syndrome (AIDS), transplant rejection, allergy, infection by a
 CC pathological agent or organism, inflammatory disorders, arthritis, a
 CC hematopoietic disorder, a skin disorder, atherosclerosis, stenosis, a
 CC neurological disease, Alzheimer's disease, trauma, spinal cord injury
 CC and skeletal disorders.

CC Sequence 95 AA:

Query Match 100.0%; Score 496; DB 21; Length 95;

Best Local Similarity 100.0%; Pred. No. 4.5e-48;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWVLMALALLHCYADSGCKLEDWEKTIINSISIPYKELLQEFIDSDAAEANG 60
 DB 1 MKLWVLMALALLHCYADSGCKLEDWEKTIINSISIPYKELLQEFIDSDAAEANG 60
 QY 61 KFKQCFLNOSHRTLNFGIMHTVYDSIWCNMKSN 95
 DB 61 KFKQCFLNOSHRTLNFGIMHTVYDSIWCNMKSN 95

RESULT 6

AAV65394

ID AAV65394 standard; Protein: 95 AA.

AC AAV65394;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1555.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.

XX Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

PF 09-APR-1999; 99WO-1B00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX MPI: 2000-038446/03.

DR N-PSDB: AAZ43008.

PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
 XX
 PS Claim 3; Page 818; 837pp; English.

CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAV64651 to
 CC AAV65438 represent the EST-related proteins corresponding to AA242265 to
 CC AA243052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be

CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AA242249 to AA242264 and AAV64644 to AAV64650 represent
 CC sequences used in the exemplification of the present invention.

CC Sequence 95 AA:

Query Match 100.0%; Score 496; DB 21; Length 95;

Best Local Similarity 100.0%; Pred. No. 4.5e-48;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWVLMALALLHCYADSGCKLEDWEKTIINSISIPYKELLQEFIDSDAAEANG 60
 DB 1 MKLWVLMALALLHCYADSGCKLEDWEKTIINSISIPYKELLQEFIDSDAAEANG 60
 QY 61 KFKQCFLNOSHRTLNFGIMHTVYDSIWCNMKSN 95
 DB 61 KFKQCFLNOSHRTLNFGIMHTVYDSIWCNMKSN 95

RESULT 7

AA31682

ID AAB31682 standard; Protein: 95 AA.

AC AAB31682;

DT 30-APR-2001 (first entry)

DE An endometrial specific steroid binding factor III.

XX Human; endometrial specific steroid binding factor; hESF; hESFII;
 KW hESFIII; inflammation; asthma; rhinitis; cystic fibrosis; airway disease;
 KW neoplasia; atopy; phospholipase A2; polychlorinated biphenyl; chemotaxis;
 KW phagocytosis; platelet aggregation; eicosanoid; endometrial cell.

XX Homo sapiens.

XX OS

XX FH

XX FT

XX FT

XX PN

XX US6174992-B1.

XX 16-JAN-2001.

PF 08-MAR-1999; 99US-0263810.

PR 21-MAR-1996; 96US-0014724.

PR 21-MAR-1997; 97US-0821451.

XX (HUMA-) HUMAN GENOME SCI INC.

XX NI J, Yu G, Gentz R;

XX MPI: 2001-158477/16.

DR N-PSDB: AAF25214.

PT New human endometrial specific steroid binding factors, useful for

PT treating and preventing inflammation, asthma, rhinitis, cystic

PT fibrosis, airway disease, neoplasia and atopy

XX Claim 1; Fig 3; 36pp; English.

CC The present sequence represents a human endometrial specific steroid
 CC binding factor (hESF). The specification describes hESF1, hESFII, and
 CC hESFIII, hESF1, II and III polypeptides, and polynucleotides encoding
 CC them are useful for treating and preventing inflammation, asthma,
 CC rhinitis, cystic fibrosis, airway disease, neoplasia and atopy,
 CC inhibiting phospholipase A2 activity, binding polychlorinated
 CC biphenyls, reducing foreign protein antigenicity, inhibiting monocyte

OY 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIMCNMKN 95
 DB 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIMCNMKN 95

RESULT 10
 AAY60038
 ID AAY60038 standard; Protein; 108 AA.

AC AAY60038:

31-JAN-2000 (first entry)

Human endometrium tumour EST encoded protein 98.

Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
 treatment; uterine; gene therapy; expressed sequence tag.

OS Homo sapiens.

PN DE19817948-A1.

PD 21-OCT-1999.

PF 17-APR-1998; 98DE-1017948.

PR 17-APR-1998; 98DE-1017948.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI: 1999-591957/51.

DR N-PSDB; AAZ42013.

New nucleic acid sequences expressed in uterine cancer tissues, and
 derived polypeptides, for treatment of uterine and endometrial cancer
 and identification of therapeutic agents -

Claim 23; Page 314; 444pp; German.

This invention describes novel human nucleic acid (cDNA) sequences (A),
 that are highly expressed in uterine tumour tissue and which have
 anticancer and cytostatic activity. (A) are used (i) for recombinant
 expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 are used (i) to identify agents suitable for treatment of uterine or
 endometrial cancer; (ii) directly for treating these forms of cancer
 (including expression from gene therapy vectors) and (iii) for
 generation of specific antibodies. (A) are identified by assembling ESTs
 (expressed sequence tags) from a particular tissue type before comparison
 of expression patterns. This allows a significantly longer fragment of
 the gene to be revealed, so should reduce the number of failures
 associated with the fact that ESTs from different libraries may represent
 different parts of the same unknown gene, distorting the estimated
 frequency of occurrence in a particular tissue. AAY59941-V60328 represent
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in AAZ41981-242121.

XX Sequence 108 AA;

Query Match 100.0%; Score 496; DB 20; Length 108;
 Best Local Similarity 100.0%; Pred. No. 5.3e-48;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLLMVLMLAALLHHCYADSGCKLLEDVVEKTIINSDISIPEYKELLOEFTDSDAAEAMG 60
 DB 14 MKLLMVLMLAALLHHCYADSGCKLLEDVVEKTIINSDISIPEYKELLOEFTDSDAAEAMG 73

OY 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIMCNMKN 95

DB 74 KFKOCFLNOSHRTLNKNGMLMHTVYDSIMCNMKN 108

RESULT 11
 AAW61649
 ID AAW61649 standard; peptide; 74 AA.

AC AAW61649:

27-OCT-1998 (first entry)

Non-ocular disease marker 3.

Human; non-ocular disease; tear; cancer; breast; prostate.

OS Homo sapiens.

PN WO9835229-A1.

PD 13-AUG-1998.

PF 06-FEB-1998; 98WO-AU00071.

PR 07-FEB-1997; 97AU-0005009.

PA (MACO-) MACQUARIE RES LTD.

PA (UNIX) UNISEARCH LTD.

PI Bolis S, Goolley AA, Herbert B, Molloy M, Morris C;

PI Walsh B, Wilcox M, Williams KL;

DR WPI: 1998-447373/38.

Screening for non-ocular disease - by analysing tears for marker

proteins, particularly indicative of cancer and genetic disease,

also new proteins and nucleic acid encoding them

Claim 6; Page 9; 14pp; English.

The markers AAW61647-W61649 are used for screening for, or detecting,
 non-ocular disease by analysing tears. Biochemicals, specifically
 proteins, are isolated from tears, particularly by chromatography or
 electrophoresis, especially two-dimensional polyacrylamide gel
 electrophoresis (2D-PAGE), then detected, e.g. with labelled specific
 reagents, in (radio)immunoassay. The method is used to detect cancer,
 particularly of breast or prostate, or a genetic disease, in humans or
 animals.

XX Sequence 74 AA;

Query Match 78.6%; Score 390; DB 19; Length 74;
 Best Local Similarity 98.6%; Pred. No. 2.6e-36;
 Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 19 DSCCKLEDMVEKTIINSDISIPEYKELLOEFTDSDAAEAMGKFKOCFLNOSHRTLNKNG 78

DB 1 DSCCKLEDMVEKTIINSDISIPEYKELLOEFTDSDAAEAMGKFKOCFLNOSHRTLNKNG 60

OY 79 LMMHTVYDSIMCNM 92

DB 61 LMMHTVYDSIMCNM 74

RESULT 12

AAU83620
 ID AAU83620 standard; Protein; 93 AA.

AC AAU83620:

08-MAY-2002 (first entry)

Human PRO protein, Seq ID No 58.

Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
 breast cancer; prostate tumour; rectal tumour; liver tumour;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 11:38:34 ; Search time 45 Seconds

(without alignments)
62.115 Million cell updates/sec

Title: US-09-806-302A-2

Perfect score: 496
Sequence: 1 MKLLVLMALALLHCYADS.....NEGLMMHTYDSIWMNKSX 95

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	496	100.0	95	3	US-08-821-451A-6	Sequence 6, App11
2	496	100.0	95	4	US-09-263-810-6	Sequence 6, App11
3	496	100.0	95	4	US-09-833-169-6	Sequence 6, App11
4	281	56.7	93	1	US-08-455-896-2	Sequence 2, App11
5	281	56.7	93	2	US-08-933-149-2	Sequence 2, App11
6	281	56.7	93	2	US-09-082-343-2	Sequence 2, App11
7	281	56.7	93	3	US-09-082-253-2	Sequence 2, App11
8	281	56.7	93	4	US-09-215-818-5	Sequence 2, App11
9	281	56.7	93	4	PCT-US96-08235-2	Sequence 5, App11
10	196.5	39.6	95	3	US-08-821-451A-27	Sequence 27, App11
11	196.5	39.6	95	4	US-09-263-810-27	Sequence 27, App11
12	196.5	39.6	95	4	US-09-583-169-27	Sequence 27, App11
13	195.5	39.4	95	1	US-08-455-896-7	Sequence 27, App11
14	195.5	39.4	95	2	US-08-933-149-7	Sequence 7, App11
15	195.5	39.4	95	2	US-09-082-343-7	Sequence 7, App11
16	195.5	39.4	95	3	US-09-082-253-7	Sequence 7, App11
17	195.5	39.4	95	5	PCT-US96-08235-7	Sequence 7, App11
18	125.5	13.1	1255	1	US-08-947-823-3	Sequence 3, App11
19	63.5	12.8	582	1	US-08-431-080-16	Sequence 16, App11
20	63.5	12.8	582	2	US-08-938-534-16	Sequence 16, App11
21	63.5	12.8	582	4	US-09-545-294-16	Sequence 16, App11
22	61.5	12.4	1604	4	US-09-004-838-95	Sequence 95, App11
23	61	12.3	446	2	US-08-922-171-3	Sequence 3, App11
24	61	12.3	472	2	US-08-922-171-2	Sequence 3, App11
25	59.5	12.0	1257	3	US-08-947-823-5	Sequence 5, App11
26	59.5	12.0	114	1	US-08-031-399-3	Sequence 3, App11
27	59.5	12.0	114	1	US-08-031-399-6	Sequence 6, App11

28	59.5	12.0	114	1	US-08-031-399-12	Sequence 12, App11
29	59.5	12.0	114	1	US-08-393-305-3	Sequence 3, App11
30	59.5	12.0	114	1	US-08-393-305-6	Sequence 6, App11
31	59.5	12.0	114	1	US-08-726-817-3	Sequence 3, App11
32	59.5	12.0	114	1	US-08-726-817-6	Sequence 6, App11
33	59.5	12.0	114	1	US-08-504-042-3	Sequence 3, App11
34	59.5	12.0	114	1	US-08-504-042-6	Sequence 6, App11
35	59.5	12.0	114	1	US-08-504-042-12	Sequence 12, App11
36	59.5	12.0	114	2	US-08-725-969-3	Sequence 3, App11
37	59.5	12.0	114	2	US-08-725-969-6	Sequence 6, App11
38	59.5	12.0	114	2	US-08-794-524-3	Sequence 3, App11
39	59.5	12.0	114	2	US-08-794-524-6	Sequence 6, App11
40	59.5	12.0	114	4	US-09-189-193-3	Sequence 3, App11
41	59.5	12.0	114	4	US-09-189-193-6	Sequence 6, App11
42	59.5	12.0	114	5	PCT-US94-03793-3	Sequence 3, App11
43	59.5	12.0	114	5	PCT-US94-03793-6	Sequence 6, App11
44	59.5	12.0	114	5	PCT-US94-03793-12	Sequence 12, App11
45	59.5	12.0	122	1	US-08-300-903A-3	Sequence 3, App11

ALIGNMENTS

```
RESULT 1
US-08-821-451A-6
: Sequence 6, Application US/08821451A
: Patent No. 6066724
:
: GENERAL INFORMATION:
: APPLICANT: JIAN NI, GUO-LIANG YU and Reiner Gentz
: TITLE OF INVENTION: Human Endometrial Specific Steroid-
: TITLE OF INVENTION: Binding Factor I, II and III
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
: ADDRESSEE: CECCHI, STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/821,451A
: FILING DATE: March 21, 1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/014,724
: FILING DATE: March 21, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: MULLINS, J.G.
: REGISTRATION NUMBER: 33,073
: REFERENCE/DOCKET NUMBER: 325800-521 (PR257)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
:
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 95 AMINO ACIDS
: TYPE: AMINO ACID
: STRANDEDNESS:
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
:
: US-08-821-451A-6
:
: Query Match 100.0%; Score 496; DB 3; Length 95;
: Best Local Similarity 100.0%; Pred. No. 7.2e-54;
: Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: OY 1 MKLLVLMALALLHCYADSGCKLLEDVMEKTIINSIDISPEYKELLQEFIDSDAAAEAMG 60
```

Db 1 MKLWMLAALLHCYADSGCKLEDVMEKTI NSDISIPEYKELLQEFIDSDAAAEAMG 60
QY 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95
Db 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95

RESULT 2

US-09-263-810-6
Sequence 6, Application US/09263810
Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARBELL, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1700
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-263-810-6
Query Match 100.0%; Score 496; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 7.2e-54;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MKLWMLAALLHCYADSGCKLEDVMEKTI NSDISIPEYKELLQEFIDSDAAAEAMG 60
QY 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95
Db 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95

Db 1 MKLWMLAALLHCYADSGCKLEDVMEKTI NSDISIPEYKELLQEFIDSDAAAEAMG 60
QY 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95
Db 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95

Db 1 MKLWMLAALLHCYADSGCKLEDVMEKTI NSDISIPEYKELLQEFIDSDAAAEAMG 60
QY 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95
Db 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95

RESULT 3

US-09-583-169-6
Sequence 6, Application US/09583169
Patent No. 6389948
GENERAL INFORMATION:
APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-

;; TITLE OF INVENTION: Binding Factor I, II and III
;; NUMBER OF SEQUENCES: 27
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CARBELL, BYRNE, BAIN, GILFILLAN,
;; ADDRESSEE: CECCHI, STEWART & OLSTEIN
;; STREET: 6 BECKER FARM ROAD
;; CITY: ROSELAND
;; STATE: NEW JERSEY
;; COUNTRY: USA
;; ZIP: 07068
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5 INCH DISKETTE
;; COMPUTER: IBM PS/2
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: WORD PERFECT 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/583,169
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/821,451
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: MULLINS, J.G.
;; REGISTRATION NUMBER: 33,073
;; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201-994-1744
;; TELEFAX: 201-994-1700
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 95 AMINO ACIDS
;; TYPE: AMINO ACID
;; STRANDEDNESS:
;; TOPOLOGY: LINEAR
;; MOLECULE TYPE: PROTEIN
;; US-09-583-169-6
Query Match 100.0%; Score 496; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 7.2e-54;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWMLAALLHCYADSGCKLEDVMEKTI NSDISIPEYKELLQEFIDSDAAAEAMG 60
Db 1 MKLWMLAALLHCYADSGCKLEDVMEKTI NSDISIPEYKELLQEFIDSDAAAEAMG 60
QY 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95
Db 61 KFKOCFLNOSHRTLKNGLMHTVYDSIMCNMKN 95

RESULT 4

US-08-455-896-2
Sequence 2, Application US/08455896
Patent No. 5668267
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25


```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/455,896
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: HOLLAND, DONALD R.
: REGISTRATION NUMBER: 35,197
: REFERENCE/DOCKET NUMBER: 952726
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (314) 727-5188
: TELEFAX: (314) 727-6092
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
:
US-08-455-896-2

Query Match
Best Local Similarity 56.7%; Score 281; DB 1; Length 93;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLIMVLMALALLHCYADSGCKLEDVMEKTIINSISIPYEKELLOEFIDSDAAAEAMG 60
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 MKLIMVLMALALSHCHAGSGCPLEENVISKTIINPOVSKTEYKELLOEFIDNATTNAID 60
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 61 KFKOCFLNOSHRTLKNFGLMHTVYDSIMCNCM 92
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 61 ELKECFLNQTFDETLNVEVFMQLIYDSSLCDL 92
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 5
US-08-933-149-2
: Sequence 2, Application US/08933149
: Patent No. 5922836
: GENERAL INFORMATION:
: APPLICANT: WATSON, MARK A.
: TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
: TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: HOWELL & HAFERKAMP, L.C.
: STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
: CITY: ST. LOUIS
: STATE: MISSOURI
: COUNTRY: USA
: ZIP: 63105-1817
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/933,149
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: HENDERSON, MELDIE W.
: REGISTRATION NUMBER: 37,848
: REFERENCE/DOCKET NUMBER: 6029-6040
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (314) 727-5188
: TELEFAX: (314) 727-6092
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
```

```

: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
:
US-08-933-149-2

Query Match
Best Local Similarity 56.7%; Score 281; DB 2; Length 93;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLIMVLMALALLHCYADSGCKLEDVMEKTIINSISIPYEKELLOEFIDSDAAAEAMG 60
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 MKLIMVLMALALSHCHAGSGCPLEENVISKTIINPOVSKTEYKELLOEFIDNATTNAID 60
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 61 KFKOCFLNOSHRTLKNFGLMHTVYDSIMCNCM 92
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 61 ELKECFLNQTFDETLNVEVFMQLIYDSSLCDL 92
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 6
US-09-082-343-2
: Sequence 2, Application US/09082343
: Patent No. 5968754
: GENERAL INFORMATION:
: APPLICANT: WATSON, MARK A.
: TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
: TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
: STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
: CITY: ST. LOUIS
: STATE: MISSOURI
: COUNTRY: USA
: ZIP: 63105-1817
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/082,343
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/455,896
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: HOLLAND, DONALD R.
: REGISTRATION NUMBER: 35,197
: REFERENCE/DOCKET NUMBER: 952726
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (314) 727-5188
: TELEFAX: (314) 727-6092
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
:
US-09-082-343-2

Query Match
Best Local Similarity 56.7%; Score 281; DB 2; Length 93;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLIMVLMALALLHCYADSGCKLEDVMEKTIINSISIPYEKELLOEFIDSDAAAEAMG 60
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 MKLIMVLMALALSHCHAGSGCPLEENVISKTIINPOVSKTEYKELLOEFIDNATTNAID 60
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 61 KFKOCFLNOSHRTLKNFGLMHTVYDSIMCNCM 92
: ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:26:19 ; Search time 57 Seconds

(without alignments)
160,224 Million cell updates/sec

Title: US-09-806-302a-2

Perfect score: 496

Sequence: 1 MKLWLMIALALHHCYADS.....NFGLMHPTVYDSIMCNKEN 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	39.4	95	1 B0R73	prostatic steroid-
2	76.5	15.4	280	2 T38449	hypothetical prote
3	76	15.3	212	2 H87530	D-alanyl-D-alanine
4	73.5	14.8	590	2 A96656	unknown protein, 3
5	71.5	14.4	131	2 F70880	hypothetical prote
6	71	14.3	303	2 E97850	cell division prot
7	70	14.1	98	1 B0R72	prostatic steroid-
8	70	14.1	303	2 B71638	cell division prot
9	69.5	14.0	412	2 T47321	hypothetical prote
10	69.5	14.0	1048	2 C96669	protein FIN19.15 l
11	68.5	13.8	173	2 T01889	hypothetical prote
12	68.5	13.8	213	2 C96925	probable membrane
13	68.5	13.8	219	2 F69435	hypothetical prote
14	68.5	13.8	496	2 T46356	hypothetical prote
15	68	13.7	516	2 H87369	tryptophan halogen
16	67.5	13.6	457	1 GMEKCC	indole-3-glycerol-
17	67	13.5	457	2 T29741	hypothetical prote
18	67	13.5	738	2 C84700	hypothetical prote
19	66.5	13.4	95	2 S68231	EHG22 protein prec
20	66.5	13.4	234	2 F64104	hypothetical prote
21	66	13.3	638	2 T47569	hypothetical prote
22	65.5	13.2	162	2 T49124	interleukin-15 - m
23	65.5	13.2	429	2 D84554	probable cyclin 2
24	65.5	13.2	603	1 W1W8	El protein - human
25	65	13.1	269	2 T04394	NBS-1RR type resis
26	65	13.1	975	2 A86258	protein F5011.4 [1
27	65	13.1	1255	2 T06267	nematodes resistan
28	65	13.1	2145	2 S61041	glutamate synthase
29	64.5	13.0	153	2 E82928	hypothetical prote

30	64.5	13.0	288	2 H84757	hypothetical prote
31	64.5	13.0	490	2 T31646	hypothetical prote
32	64.5	13.0	610	2 S59558	dynamitin-like prote
33	64	12.9	629	2 A96657	unknown protein, 7
34	63.5	12.8	264	2 G69278	glutamine ABC tran
35	63.5	12.8	269	2 G75148	hypothetical prote
36	63.5	12.8	582	2 S69720	hypothetical prote
37	63.5	12.8	1294	2 S58307	hypothetical 149.2
38	63	12.7	131	2 D81908	hypothetical prote
39	63	12.7	288	2 AB0092	probable flagellar
40	63	12.7	785	2 D71653	cell surface antiq
41	62.5	12.6	91	1 UGRB	uteroglobin precu
42	62.5	12.6	113	2 JC4143	molt-inhibiting ho
43	62.5	12.6	137	2 F97763	hypothetical prote
44	62.5	12.6	223	2 S50404	hypothetical prote
45	62.5	12.6	429	2 T04667	cyclin 2b - Arabid

ALIGNMENTS

RESULT 1

B0R73
prostatic steroid-binding protein chain C3 precursor - rat
N:Alternate names: prostatein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Dec-1982 #sequence_revision 14-Nov-1983 #text_change 28-May-1999
C:Accession: A92395; A92433; A91108; A42392; A03250
C:Paraker, M.G.; White, R.; Hurst, H.; Needham, M.; Tilly, R.
J. Biol. Chem. 258, 12-15, 1983
A>Title: Prostatic steroid-binding protein. Isolation and characterization of C3 gene
A:Reference number: A92395; MIM:83082848; PMID:6294095
A:Accession: A92395
A:Molecule type: mRNA
A:Residues: 1-95 <PAR>
A:Cross-references: GB:V01263; GB:J00777; MID:956993; PIDN:CA24577.1; PID:956994
R:Vyskocil, D.H.; Perry, S.T.; Lea, O.A.; Stafford, D.W.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 258, 8861-8866, 1983
A>Title: Isolation of two genomic sequences encoding the M-r = 14,000 subunit of rat
A:Reference number: A92433; MIM:8328526; PMID:6190812
A:Accession: A92433
A:Molecule type: mRNA
A:Residues: 1-78, 'S', '80-95 <VIS>
R:Peeters, B.; Rombaux, W.; Mous, J.; Heyns, W.
Eur. J. Biochem. 115, 115-121, 1981
A>Title: Structural studies on rat prostatic binding protein. The primary structure o
A:Reference number: A91108; MIM:81188769; PMID:7014218
A:Accession: A91108
A:Molecule type: protein
A:Residues: 19-95 <PEP>
R:Ran, J.A.; Marschke, K.B.; Ho, K.C.; Perry, S.T.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 267, 4456-4466, 1992
A>Title: Response elements of the androgen-regulated C3 gene.
A:Reference number: A42392; MIM:92165796; PMID:1537831
A:Accession: A42392
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-16, 'T', '18-95 <TAN>
A>Note: sequence inconsistent with the nucleotide translation
C:Note: sequence extracted from NCBI backbone (NCBIN:83851, NCBI:83856)
C:Comment: C3 is encoded by two unique genes that differ from each other only in the
C:Comment: Steroid-binding protein, the principal androgen-dependent secretory protei
ng C2 and C3 chains. The chains of each dimer are linked by disulfide bonds. This pro
C:Superfamily: uteroglobin
C:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-95/Product: prostatic steroid-binding protein C3 chain #status experimental
F:35/Binding site: carbohydrate (Asn) (covalent) #status experimental
Query Match 39.4%; Score 195.5; DB 1; Length 95;
Best local Similarity 35.8%; Pred. No. 3,9e-13;
Matches 34; Conservative 30; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLMVLMAALLHCVAD-SGCKLLEDVMEKTIINSISIPYKELLOEFDSDAAEAAM 59
 DB 1 MKLVLEFLVLTIPICCAAGSGSILDEVRGNSVTIHLMDLKVPRVYODHFEKAV 60

QY 60 GKFKOCFLNDSHRTLNFGMLMHTVYDSIWCNKS 94
 DB 61 KOFKOCFLDQTDKTLVGVYMAEAFNFSQCOOPS 95

RESULT 2

hypothetical protein SPAC7E2.02 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T38449
 R:Murphy, L.; Harrell, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 Submitted to the EMBL Data Library, September 1997
 A:Reference number: 121749
 A:Accession: T38449
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-280 <MUR>
 A:Cross-references: EMBL:P98978; PIDN:CAB11676.1; GSPDB:GN00066; SPDB:SPAC7E2.02
 A:Experimental source: strain 972h-; cosmid c27E2
 C:Genetics:
 A:Gene: SPDB:SPAC7E2.02
 A:Map position: 1
 A:Introns: 185/3; 225/3

Query Match 15.4%; Score 76.5; DB 2; Length 280;
 Best Local Similarity 30.3%; Pred. No. 2.3;
 Matches 23; Conservative 10; Mismatches 28; Indels 15; Gaps 3;

QY 24 LLEDVMEKTIINSISIPYKELLOEFDSD--AAAFAMKRFQ-----CFLNQSHRT 73
 DB 68 LADVLLSVYTGVCVFSYMDLKLKELVDIDAQAAERESKLOESDKETPVMLNNSHYV 127
 QY 74 LKNFGMLMHTVYDSIW 89
 DB 128 AKT-----PEIQDEPW 138

RESULT 3

H87530
 D-alanyl-D-alanine dipeptidase [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: H87530

R:Metzger, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete genome sequence of Caulobacter crescentus.
 A:Reference number: AB7249; MUID:2113698; PMID:11259647

A:Accession: H87530
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-212 <STO>
 A:Cross-references: GB:AE005673; NID:g13423786; PIDN:AAK2424.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2273
 C:Superfamily: Enterococcus faecium transposon Tn1546 D-alanine-D-alanine dipeptidase

Query Match 15.3%; Score 76; DB 2; Length 212;
 Best Local Similarity 29.9%; Pred. No. 1.9;
 Matches 26; Conservative 13; Mismatches 16; Indels 32; Gaps 5;

QY 18 ADGCKLLEMEVET-----INSDISPEYKE--LLOEFDSDAAEAAMG 60
 DB 19 AETGDKLASDLVDVTTTPGKIDIRAGANNMGIDPLYSAAVYLOR-----PAAEALG 73
 QY 61 KFKOCFLNDSHRTL--KNFGMLMHTVY 85

DB 74 RI-----HRLAAKGYGLLTHDAY 92

RESULT 4
 A96656
 unknown protein, 38394-36551 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: A96656
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizart, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: A96656
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-590 <STO>
 A:Cross-references: GB:AE005173; NID:g10092251; PIDN:AGL12665.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: p16M19.15
 A:Map position: 1

Query Match 14.8%; Score 73.5; DB 2; Length 590;
 Best Local Similarity 32.3%; Pred. No. 11;
 Matches 21; Conservative 8; Mismatches 23; Indels 13; Gaps 3;

QY 13 LLMCYDSC-----CULEDMVEKTIINSISIPYKELLOEFDSDAAEA--MKG 61
 DB 291 LISCLNNGRMSDASRLSDMLEKINIPDLVF--FNALIDAFVKECKLYEAKLYDEMYK 348
 QY 62 FKOCF 66
 DB 349 SKHCF 353

RESULT 5

F70880
 hypothetical protein RV2759C - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: F70880

R:Coyle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Genies, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: F70880
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-131 <COL>
 A:Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA15555.1; PID:ell17
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV2759C

Query Match 14.4%; Score 71.5; DB 2; Length 131;
 Best Local Similarity 21.2%; Pred. No. 3.3;
 Matches 21; Conservative 18; Mismatches 35; Indels 25; Gaps 1;

QY 5 MYMLMAALLHCVADSGCKLLEDVMEKTIINSISIPYKELLO----- 47
 DB 1 MIVDTSAIVAIVSGEQAQVLEALERSPNSRMSAPNVYELCAIMQRDRPEISRLVDRL 60

QY 48 -----EFIDSDAAEAMGKFKOCFLNOSHRTLNFG 78
B71638
Db 61 LDDYGIQVEAVDADQARVAADAYRGRSGCHPARLNIG 99

RESULT 6

cell division protein ftsy [Imported] - Rickettsia conorii (strain/Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
R:Accession: E97850
R:Orgate, H.; Audic, S.; Renezo-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; Ro-
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: E97850
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <R>
A:Cross-references: GB:AE006914; PIDN:AAL03743.1; PID:915620335; GSPDB:GN00173
C:Genetics:
A:Gene: ftsy
C:Superfamily: docking protein

Query Match 14.3%; Score 71; DB 2; Length 303;
Best Local Similarity 33.3%; Pred. No. 9.2;
Matches 20; Conservative 13; Mismatches 17; Indels 10; Gaps 2;

QY 19 DSGCKLEDAVEKINDISPEKELDF-----IDSDAAAEAMGKFKOCFLNOS 70
B71638
Db 33 DAG--TLEELLELLISDSISVYTHIEEFKNVKFDKTDSPVKEALKEIQOLSKS 90

RESULT 7

prostatic steroid-binding protein chain C2 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 24-Sep-1999
R:Accession: A03251; A26671
R:Parker, M.; Needham, M.; White, R.
Nature 298, 97-94, 1982
A:Title: Prostatic steroid binding protein: gene duplication and steroid binding.
A:Reference number: A93286; MUID:82220075; PMID:6896362
A:Accession: A03251
A:Molecule type: mRNA
A:Residues: 1-98 <PAR>
A:Cross-references: GB:J00776; NID:q206448; PIDN:AAA51641.1; PID:q206450
R:Deleay, B.; Ditsch, L.; Decourt, J.L.; Claessens, F.; Peeters, B.; Rombaux, W.
Nucleic Acids Res. 15, 1627-1641, 1987
A:Title: Rat prostatic binding protein: the complete sequence of the C2 gene and its flanking regions.
A:Reference number: A26671; MUID:8716484; PMID:2881277
A:Accession: A26671
A:Molecule type: DNA
A:Residues: 1-25, 'Q', 26-86, 'T', 88-94, 'WLMQINPGRWFSEIN'
A:Cross-references: GB:X05034; NID:956857; PIDN:CAA28708.1; PID:956858
C:Comment: Steroid-binding protein, the principal secretory protein in rat prostatic fluid.
The chains of each dimer are linked by disulfide bonds.
C:Superfamily: uteroglobin
C:Keywords: heterotetramer; prostate; steroid binding
F:120/Domain: signal sequence #status predicted <SIG>
F:21-98/Product: prostatic steroid-binding protein chain C2 #status predicted <MAT>

Query Match 14.1%; Score 70; DB 1; Length 98;
Best Local Similarity 27.3%; Pred. No. 3.4;
Matches 27; Conservative 20; Mismatches 36; Indels 16; Gaps 5;

QY 4 LMTLMLALLHCYADSG-----CKLEDM-VEKINSISPEYKELLOEFTSDAAA 56
B71638
Db 3 LSLCILTLIVCCYEAAGQTLGACQALQDVTITFLNPF---BELKRELEEFAPPEAV 59

QY 57 EAMGKFKOCFLNOSHRTLNFGFLMHT--VYDSIWCNKK 93
B71638
Db 60 EANLKVRCI---NKLIMYGDLSMGTSLVFTMLCKDVK 94

RESULT 8
B71638
cell division protein ftsy (fsy) RP775 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
R:Accession: B71638
R:Andersen, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9623893
A:Accession: B71638
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-303 <AND>
A:Cross-references: GB:AJ235273; GB:AJ235269; NID:93861237; PIDN:CAI5202.1; PID:9386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: ftsy; RP775
C:Superfamily: docking protein

Query Match 14.1%; Score 70; DB 2; Length 303;
Best Local Similarity 35.2%; Pred. No. 12;
Matches 19; Conservative 10; Mismatches 17; Indels 8; Gaps 1;

QY 25 LEDVVEKINSISPEYKELDF-----IDSDAAAEAMGKFKOCFLNOS 70
B71638
Db 37 LNELEELLISDSISVYTHIEEFKNVKFDKTDSPVKEALKEIQOLSKS 90

RESULT 9

hypothetical protein T12K4_80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
R:Accession: T47321
R:Monfort, A.; Casacuberta, E.; Pulgomech, P.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
Submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24460
A:Accession: T47321
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <MON>
A:Cross-references: EMBL:AL136640
A:Experimental source: cultivar Columbia; BAC clone T12K4
C:Genetics:
A:Map position: 3
A:Introns: 30/3
A:Note: T12K4_80

Query Match 14.0%; Score 69.5; DB 2; Length 412;
Best Local Similarity 27.5%; Pred. No. 18;
Matches 25; Conservative 11; Mismatches 40; Indels 15; Gaps 3;

QY 13 LHCYADSGCKL---EDVVEKINSISPEYKELLOEFTD---SDAAAEAMGKFKOCF 66
B71638
Db 15 LLSCTYDSAPSIKKESKLSRDFSKRIQIVDAPLVQTLSSQRRLPVAHFIQTKSVN 74

QY 67 LNSHRTL-----KNFGIMMHTVYDSI 88
B71638
Db 75 LLPNYRTICALMLCFEAGFVLRARTIMDEI 105

RESULT 10

C96669
protein F1N19_15 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Accession: C96669
R:Rothfels, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Greasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marshall, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.;
A:Authors: Selzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: C96669

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1048 <STC>

A:Cross-references: GB:AE005173; NID:g6633829; PIDN:AAF19688.1; GSPDB:GN00141

C:Genetics:

A:Gene: FIN19.15

A:Map position: 1

Query Match

Best Local Similarity 14.0%; Score 69.5; DB 2; Length 1048;

Matches 22; Conservative 20; Mismatches 15; Indels 37; Gaps 6;

QY 16 CVA---DSCGKLEDMVEKTIINSIDIS---SIPE---YKLLDEPIDSDA 54

DB 222 CYSGRNSDAARMRLDMKKRSINPDVYTFALIDVFKQGNLDEAQLYKEMIOSSVDPN- 280

QY 55 AAEAMGKFKOCFLNOSHRTLNKFGIMMH-TVYDS 87

DB 281 -----NVTYSIN-GLCHMGRLYDA 300

RESULT 11

T01889

hypothetical protein F8M12.4 - Arabidopsis thaliana

N:Alternate names: hypothetical protein F25124.80

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jul-2000

C:Accession: T01889; T04282

R:Madan, C.; Graves, T.; Cotton, M.; Modde, T.

Submitted to the EMBL Data Library, July 1998

A:Description: The sequence of A. thaliana F8M12.

A:Reference number: Z14450

A:Accession: T01889

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-173 <MAD>

A:Cross-references: EMBL:AF080118; NID:g3513725; PID:g3513722

A:Experimental source: cultivar Columbia

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15261

A:Accession: T04282

A:Molecule type: DNA

A:Residues: 1-173 <BEV>

A:Cross-references: EMBL:AL049525

A:Experimental source: cultivar Columbia; BAC clone F25124

C:Genetics:

A:Map position: 4

A:Note: F8M12.4; F25124.80

C:superfamily: Arabidopsis thaliana hypothetical protein F8M12.2

Query Match

Best Local Similarity 13.8%; Score 68.5; DB 2; Length 173;

Matches 27; Conservative 17; Mismatches 36; Indels 31; Gaps 6;

QY 11 ALLHCVADSGC-----KLEDMV-----EKTINSDISPEYKELLOEFIDSDAAE 57

DB 2 APLLHRYSNVTCGKRVTVRCDDKIYDTKDGESRTINEEINSGPYKRIYDFPDEDOGYL 61

QY 58 AMGKFKOCFLNOSHRTLNKFGIMMH-TVYD-----STMCNM 92

DB 62 LNTDGYVLMNDNRG--KNGIODLVHTATADDINDIYDEMKIWCET 110

RESULT 12

C96925

probable membrane protein, CF-20 family [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: C96925

R:Noll, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, O.; Gibson, R.; L

.. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: C96925

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-213 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78190.1; PID:g15023041; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0208

Query Match

Best Local Similarity 13.8%; Score 68.5; DB 2; Length 213;

Matches 26; Conservative 18; Mismatches 31; Indels 33; Gaps 4;

QY 3 LNLVLMALLLHCY---ADSGKLEDMVEKTIINSIDIS--IPEYKEL----- 45

DB 11 LMLLLVSTWLMGCKRKVPDSEAKIISDFVVKGDKTKISQIIKDKELDKYKREKDF 70

QY 46 LQEFIDSDAAAEAMGKFKOCFLNOSHRTLNKFGIMMH-TVYDSIWCNMK 93

DB 71 VKELKSSSSSGDLQITDAOC-----DEYDVSVALK 102

RESULT 13

F69435

hypothetical protein AF1487 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: F69435

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

.. Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Goodyear, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Overbeek, T.; Cotton, M.D.; Spijgs, T.; Artlisch, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Moese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: F69435

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-219 <KLE>

A:Cross-references: GB:AE001000; GB:AE000782; NID:g2689323; PIDN:AA89768.1; PID:g264

Query Match

Best Local Similarity 13.8%; Score 68.5; DB 2; Length 219;

Matches 24; Conservative 17; Mismatches 36; Indels 21; Gaps 5;

QY 1 MKLLVLMV---AALLHCVADSGCKLEDMVEKTIINSIDISPEYKELLOEFIDSD 51

DB 4 MRRMLTLFLVGLAVVL-----SGCATLS--VSKYKNDGSESVKLVINTSSFYVGLA 56

QY 52 SDAAEAMGKFKOCFLNOSHRTLNKFGIMMH-TVYDSIM 89

DB 57 EGAKKGGYESTLRSPLEIPEEMRD-----KVSYDEVW 89

RESULT 14

T46356

hypothetical protein DKFZp434G2016.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46356

R:Koehner, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:39:00 ; Search time 20 Seconds
(without alignments)
456.639 Million cell updates/sec

Title: US-09-806-302a-2

Perfect score: 95

Sequence: 1 MKLWVLMIALILHICVADS.....NFGLMHPTVYDSIMCKSN 95

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9	9.5	705	2	S45769	probable membrane
2	8	8.4	249	2	T12463	hypothetical prote
3	8	8.4	256	2	H75404	outer membrane pro
4	8	8.4	552	1	E57987	cytochrome c-type
5	8	8.4	552	2	D86101	hypothetical prote
6	8	8.4	552	2	H91260	cytochrome c-type
7	8	8.4	937	2	P00864	hypothetical prote
8	8	8.4	2105	1	A44059	genome polypeptid
9	7	7.4	76	2	T03860	TA20 protein - com
10	7	7.4	107	2	E82494	conserved hypothet
11	7	7.4	239	2	A97358	glucose-inhibited
12	7	7.4	295	2	AB3613	glucose-1-phosphat
13	7	7.4	354	2	T18464	hypothetical prote
14	7	7.4	377	1	H64114	hypothetical prote
15	7	7.4	402	2	T15677	chorismate mutase
16	7	7.4	425	2	S58672	hypothetical prote
17	7	7.4	433	2	S76553	extra sex combs pr
18	7	7.4	572	2	T45139	hypothetical prote
19	7	7.4	576	2	T36729	chaperone protein
20	7	7.4	605	2	G95853	probable serine/th
21	7	7.4	627	2	H86254	probable pyruvate
22	7	7.4	665	2	F95053	hypothetical prote
23	7	7.4	693	2	F96557	cell wall surface
24	7	7.4	758	2	B83319	probable protein k
25	7	7.4	775	2	D86261	ATP-binding protei
26	7	7.4	994	2	A70776	hypothetical prote
27	6	6.3	46	2	T46224	probable gln - My
28	6	6.3	49	2	A81605	hypothetical prote
29	6	6.3	57	2	D97724	hypothetical prote

30	6	6.3	67	2	F90882	hypothetical prote
31	6	6.3	67	2	A85736	hypothetical prote
32	6	6.3	67	2	D64894	hypothetical prote
33	6	6.3	68	2	B84265	hypothetical prote
34	6	6.3	75	2	A55419	dopamine D3 recept
35	6	6.3	79	2	B30924	hypothetical 9.8k
36	6	6.3	81	2	S71962	trefoil factor p52
37	6	6.3	81	2	C86710	hypothetical prote
38	6	6.3	89	2	S08056	hypothetical prote
39	6	6.3	92	2	T09146	late-embryogenesis
40	6	6.3	95	1	B0RT3	prostatic steroid-
41	6	6.3	95	2	G71630	integration host f
42	6	6.3	95	2	T09264	embryonic abundant
43	6	6.3	95	2	H97835	integration host f
44	6	6.3	102	2	D64363	ribosomal protein
45	6	6.3	103	2	E40899	hypothetical prote

ALIGNMENTS

RESULT 1
S45769
Probable membrane protein YBL035c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBL0414
C:Species: Saccharomyces cerevisiae
C>Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
C:Accession: S45769; S46567
R:Goffeau, A.; Joniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, submitted to the Protein Sequence Database, August 1994
A:Reference number: S45745
A:Accession: S45769
A:Molecule type: DNA
A:Residues: 1-705 <GOF>
A:Cross-references: EMBL:Z5796; NID:9536044; PID:9536045; MIPS:YBL035c
A:Experimental source: strain S288C
R:Skala, J.; Van Dyck, L.; Purnelle, B.; Goffeau, A.
Yeast 10(Suppl.A), S13-S24, 1994
A>Title: The sequence of an 8.8 kb segment on the left arm of chromosome II from Sacc and bacterial GTP cyclolhydrolase II.
A:Reference number: S46565; MUID:94378718; PMID:8091857
A:Accession: S46567
A:Molecule type: DNA
A:Residues: 1-705 <SKA>
A:Cross-references: EMBL:X74738; NID:9511140; PID:CAAS2761.1; PID:9511143
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:POL12
A:Cross-references: SGD:S0000131; MIPS:YBL035c
A:Map position: 2L
C:Keywords: transmembrane protein
F:428-444/Domain: transmembrane #status predicted <TM>

Query Match 9.5%; Score 9; DB 2; Length 705;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 44 ELLOEFTDS 52
411 ELLOEFTDS 419

RESULT 2
T12463
Hypothetical protein DKFZP564I052.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C:Accession: T12463
R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17522
A:Accession: T12463
A>Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-249 <DUE>
A:Cross-references: EMBL:AL080063
A:Experimental source: fetal brain; clone DKFZp5641052
C:Genetics:
A:Note: DKFZp5641052.1

Query Match 8.4%: Score 8; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 40 PEYKELLQ 47
 |||||
DB 136 PEYKELLQ 143

RESULT 3
H75404
outer membrane protein - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75404
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zaleski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; M0ID:20036896; PMID:10567266
A:Accession: H75404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <WHI>
A:Cross-references: GB:AE001982; GB:AE005113; NID:g6459109; PIDN:AAF10931.1; PID:g645911
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1359
A:Map position: 1
C:Superfamily: lipoprotein-28

Query Match 8.4%: Score 8; DB 2; Length 256;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 LMLALLL 14
 |||||
DB 5 LMLALLL 12

RESULT 4
E57987
cytochrome c-type biogenesis protein nrfE - *Escherichia coli* (strain K-12)
C:Species: *Escherichia coli*
C:Date: 26-Jul-1996 #sequence_revision 31-Oct-1997 #text_change 01-Mar-2002
C:Accession: A65216; E57987
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; M0ID:97426617; PMID:9278503
A:Accession: A65216
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-552 <BLAT>
A:Cross-references: GB:AE000481; GB:U00096; NID:g2367346; PIDN:AA013457.1; PID:g1790511;
A:Experimental source: strain K-12, substrain MG1655
R:Hussain, H.; Grove, J.; Griffiths, L.; Busby, S.; Cole, J.
Mol. Microbiol. 12, 153-163, 1994
A:Title: A seven-gene operon essential for formate-dependent nitrite reduction to ammoni
A:Reference number: A57987; M0ID:94335626; PMID:8057835
A:Accession: E57987
A:Molecule type: DNA
A:Residues: 1-31, 'V', 'A', '39-189, 'S', '191-217, 'L', '219, 'RLVVLGS', '227-454, 'A', '456-480,
A:Cross-references: EMBL:X72298; NID:g404302; PIDN:CAA51045.1; PID:g581146

C:Genetics:
A:Gene: nrfE
A:Note: this is one of the seven gene operon (nrf) essential for formate-dependent n
C:Function:
A:Description: an inner membrane protein, possible component of a heme lyase, which i
ent nitrite reductase activity
C:Superfamily: nrfE protein
C:Keywords: membrane protein

Query Match 8.4%: Score 8; DB 1; Length 552;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 LMLALLL 14
 |||||
DB 114 LMLALLL 121

RESULT 5
D86101
hypothetical protein nrfE [imported] - *Escherichia coli* (strain O157:H7, substrain ED
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: D86101
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; M0ID:21074935; PMID:11206551
A:Accession: D86101
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <STO>
A:Cross-references: GB:AE005174; NID:g12519033; PIDN:AAG59272.1; GSPDB:GN00145; UNGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: nrfE
C:Superfamily: nrfE protein

Query Match 8.4%: Score 8; DB 2; Length 552;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 LMLALLL 14
 |||||
DB 114 LMLALLL 121

RESULT 6
H91260
cytochrome c-type biogenesis protein NrfE [imported] - *Escherichia coli* (strain O157:
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
C:Accession: H91260
R:Ishiyoshi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and g
A:Reference number: A99629; M0ID:21156231; PMID:11258796
A:Accession: H91260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <NAV>
A:Cross-references: GB:BA000007; PIDN:BA838479.1; PID:g13364533; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS5056
C:Superfamily: nrfE protein

Query Match 8.4%: Score 8; DB 2; Length 552;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LMLALLL 14
|||||
Db 114 LMLALLL 121

RESULT 7
P00864
hypothetical protein 937 - citrus tatter leaf virus (fragment)

N:Alternate names: ORF1

C:Species: citrus tatter leaf virus

C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 20-Jun-2000

C:Accession: P00864

R:Yoshikawa, N.; Imaizumi, M.; Takahashi, T.; Inouye, N.

J. Gen. Virol. 74, 2743-2747, 1993

A:Title: Striking similarities between the nucleotide sequence and genome organization

A:Reference number: P00864; M0ID:94103780; PMID:8277280

A:Accession: P00864

A:Molecule type: mRNA

A:Residues: 1-937 <YOS>

A:Cross-references: DDBJ:D16368; NID:g464150; PIDN:BAA03869.1; PID:g563955

A:Experimental source: strain Li-23

C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

C:Keywords: ATP

Query Match 8.4%; Score 8; DB 2; Length 937;

Best Local Similarity 100.0%; Pred. No. 6.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KELLQEFI 50
|||||

Db 55 KELLQEFI 62

RESULT 8

A44059
genome polyprotein - apple stem grooving virus (strain P-209)

N:Alternate names: RNA nucleotidyltransferase (RNA-directed); RNA replicase

N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C:Species: apple stem grooving virus; ASGV

C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 19-Jan-2001

C:Accession: A44059

R:Yoshikawa, N.; Sasaki, E.; Kato, M.; Takahashi, T.

Virol. 191, 98-105, 1992

A:Title: The nucleotide sequence of apple stem grooving capillovirus genome.

A:Reference number: A44059; M0ID:93033164; PMID:1413530

A:Accession: A44059

A:Molecule type: genomic RNA

A:Residues: 1-2105 <YOS>

A:Cross-references: GB:D14995; NID:g303496; PIDN:BAA03639.1; PID:g285608

C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

C:Keywords: ATP; nucleotide binding; nucleotidyltransferase; P-loop; RNA biosynthesis;

F:781-88/Region: nucleotide-binding motif A (P-loop)

F:843-48/Region: nucleotide-binding motif B

F:787/Binding site: ATP (Lys) #status predicted

Query Match 8.4%; Score 8; DB 1; Length 2105;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KELLQEFI 50
|||||

Db 1223 KELLQEFI 1230

RESULT 9

T03860
TA20 protein - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999

C:Accession: T03860

R:Beats, T.P.; Goldberg, R.B.

submitted to the EMBL Data Library, October 1996

A:Description: Nicotiana tabacum gene expressed in anthr.

A:Reference number: Z15122

A:Accession: T03860

A>Status: Preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-76 <BEA>

A:Cross-references: EMBL:U73164; NID:g1657813; PIDN:AAB18190.1; PID:g1657814

A:Experimental source: tissue-type anthr

C:Genetics:

A:Gene: TA20

Query Match 7.4%; Score 7; DB 2; Length 76;

Best Local Similarity 100.0%; Pred. No. 8.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LMYLMLA 10
|||||

Db 19 LMYLMLA 25

RESULT 10

E82494

conserved hypothetical protein VCA0152 [imported] - Vibrio cholerae (strain N16961 se

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 16-Feb-2001

C:Accession: E82494

R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; M0ID:20406833; PMID:10952301

A:Accession: E82494

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-107 <HELI>

A:Cross-references: GB:AE004356; GB:AE003853; NID:g9657536; PIDN:AAF96065.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0152

A:Map position: 2

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1428

Query Match 7.4%; Score 7; DB 2; Length 107;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LMAALLL 14
|||||

Db 3 LMAALLL 9

RESULT 11

A97358

glucose-inhibited division protein, GIDB (S-adenosylmethionine-dependent methyltransf

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: A97358

R:Nolling, J.; Breton, G.; Omeletchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; M0ID:21359325; PMID:21359325

A:Accession: A97358

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-239 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81652.1; PID:g15026839; GSPDB:GN00168

C:Genetics:

A:Gene: CAC3732

Query Match 7.4%; Score 7; DB 2; Length 239;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 YKEL06 48
|||||||

DB 29 YKEL06 35

RESULT 12

AB3613
glucose-1-phosphate cytidylyltransferase (EC 2.7.7.33) [imported] - *Brucella melitensis*
C:Species: *Brucella melitensis*
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AB3613
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, .; Mazur, M.; Goldstein, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A:Reference number: AD2252; PMID:11756688
A:Accession: AB3613
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54069.1; PID:q17965025; GSPDB:GN00191
C:Genetics:
A:Gene: BME110827
A:Map position: 11
C:Superfamily: glucose-1-phosphate cytidylyltransferase
C:Keywords: nucleotidyltransferase

Query Match
Best Local Similarity 7.4%; Score 7; DB 2; Length 295;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 LEDMVK 31
|||||||

DB 268 LEDMVK 274

RESULT 13

T18464
hypothetical protein C0480c - malaria parasite (*Plasmodium falciparum*)
C:Species: *Plasmodium falciparum*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C:Accession: T18464
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18937
A:Accession: T18464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-354 <LAW>
A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CA15607.1
C:Genetics:
A:Map position: 3
A:Note: C0480c

Query Match
Best Local Similarity 7.4%; Score 7; DB 2; Length 354;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 LOEFIDS 52
|||||||

DB 268 LOEFIDS 274

RESULT 14

H64114
chorismate mutase (EC 5.4.99.5) / prephenate dehydrogenase (EC 1.3.1.12) H11290 [similar
N:Contains: chorismate mutase T; prephenate dehydrogenase
C:Species: *Haemophilus influenzae*
C>Date: 18-Aug-1995 #sequence_revision 10-May-1996 #text_change 06-Oct-2000
C:Accession: H64114; T09419

R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, .; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, .; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter A>Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: H64114
A:Molecule type: DNA
A:Residues: 1-377 <TIG>

A:Cross-references: GB:U032809; GB:LA2023; NID:g3212217; PIDN:AAC22939.1; PID:q1574749
A:Note: named as homolog to a protein from *Erwinia herbicola*
R:White, O.; Clayton, R.A.; Kerlavage, A.R.; Feilschmann, R.D.; Peterson, J.; Hickey, submitted to the EMBL Data Library, May 1998
A:Reference number: Z16667
A:Accession: T09419

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-377 <WHI>
A:Cross-references: EMBL:U032809; NID:g3212217; PIDN:AAC22939.1; PID:q1574749
C:Genetics:
A:Gene: H11290
A:Start codon: GTG
C:Superfamily: tyra bifunctional enzyme
C:Keywords: intramolecular transferase; isomerase; multifunctional enzyme; oxidoreduc

Query Match
Best Local Similarity 7.4%; Score 7; DB 1; Length 377;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 KTINSDI 37
|||||||

DB 94 KTINSDI 100

RESULT 15

T15677
hypothetical protein C28C12.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15677
R:Miller, N.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of *C. elegans* cosmid C28C12.
A:Reference number: Z18387
A:Accession: T15677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-402 <MIT>
A:Cross-references: EMBL:U40797; NID:g1065916; PID:g1065921; PIDN:AAB37548.1; GSPDB:G
A:Experimental source: strain Bristol N2; clone C28C12
C:Genetics:
A:Gene: CESP:C28C12.5
A:Map position: 4
A:Introns: 29/3; 82/3; 124/3; 151/3; 258/3; 318/1; 372/3; 400/3

Query Match
Best Local Similarity 7.4%; Score 7; DB 2; Length 402;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 LLOEFTD 51
|||||||

DB 250 LLOEFTD 256

Search completed: January 25, 2003, 13:41:50
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:34:15 : Search time 10 Seconds
(without alignments)
191.696 Million cell updates/sec

Title: US-09-806-302A-2

Perfect score: 95

Sequence: 1 MKLLMVLMLAALLHCYADS.....NFGIMHTYDSITCMKSN 95

Scoring table: OLIGO
Gapop 60.0 , Capext 60.0

Searched: 122226 seqs, 20178551 residues

Word size : 0

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications-AA:*
1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppa/PCR_US07_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	95	10	US-09-110-716-31 Sequence 31, Appl
2	95	100.0	95	10	US-09-985-911-6 Sequence 6, Appl
3	77	81.1	77	10	US-09-110-716-13 Sequence 13, Appl
4	65	68.4	76	10	US-09-110-716-40 Sequence 40, Appl
5	12	12.6	13	10	US-09-757-417-29 Sequence 29, Appl
6	12	12.6	93	10	US-09-757-417-27 Sequence 27, Appl
7	12	12.6	93	10	US-09-934-054-3 Sequence 3, Appl
8	12	12.6	93	10	US-09-934-054-10 Sequence 10, Appl
9	12	12.6	93	12	US-10-007-805-503 Sequence 503, Appl
10	12	12.6	132	10	US-09-757-417-47 Sequence 47, Appl
11	12	12.6	410	12	US-10-007-805-495 Sequence 495, Appl
12	12	12.6	743	12	US-10-007-805-494 Sequence 494, Appl
13	11	11.6	1095	12	US-10-007-805-493 Sequence 493, Appl
14	11	11.6	20	10	US-09-757-417-4 Sequence 4, Appl
15	11	11.6	20	10	US-09-757-417-33 Sequence 33, Appl
16	11	11.6	20	12	US-10-007-805-499 Sequence 499, Appl
17	11	11.6	75	10	US-09-110-716-41 Sequence 41, Appl
18	10	10.5	10	10	US-09-757-417-42 Sequence 42, Appl
19	10	10.5	20	10	US-09-757-417-32 Sequence 32, Appl

20	9	9.5	9	10	US-09-757-417-5	Sequence 5, Appl
21	9	9.5	9	10	US-09-757-417-37	Sequence 37, Appl
22	9	9.5	9	10	US-09-757-417-38	Sequence 38, Appl
23	9	9.5	9	10	US-09-757-417-39	Sequence 39, Appl
24	9	9.5	9	12	US-10-007-805-500	Sequence 500, Appl
25	8	8.4	20	10	US-09-757-417-12	Sequence 12, Appl
26	8	8.4	20	10	US-09-757-417-14	Sequence 14, Appl
27	8	8.4	20	10	US-09-757-417-15	Sequence 15, Appl
28	8	8.4	20	10	US-09-757-417-17	Sequence 17, Appl
29	7	7.4	45	10	US-09-757-417-45	Sequence 45, Appl
30	7	7.4	45	10	US-09-864-761-45420	Sequence 45420, A
31	7	7.4	112	10	US-09-682-706-2	Sequence 2, Appl
32	7	7.4	112	10	US-09-919-473-2	Sequence 2, Appl
33	6	6.3	9	10	US-09-244-694-184	Sequence 184, App
34	6	6.3	22	9	US-09-865-989-11	Sequence 11, Appl
35	6	6.3	22	9	US-09-865-989-17	Sequence 17, Appl
36	6	6.3	22	9	US-09-865-989-23	Sequence 23, Appl
37	6	6.3	22	9	US-09-865-989-36	Sequence 36, Appl
38	6	6.3	22	9	US-09-865-989-87	Sequence 87, Appl
39	6	6.3	31	10	US-09-732-091-13	Sequence 13, Appl
40	6	6.3	48	10	US-09-925-299-1468	Sequence 1468, Ap
41	6	6.3	48	10	US-09-864-761-40519	Sequence 40519, A
42	6	6.3	63	10	US-09-864-761-34909	Sequence 34909, A
43	6	6.3	70	10	US-09-289-346A-1	Sequence 1, Appl
44	6	6.3	70	10	US-09-289-346A-2	Sequence 2, Appl
45	6	6.3	70	10	US-09-289-346A-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-09-110-716-31
; Sequence 31, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehner, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596.00
; CURRENT APPLICATION NUMBER: US/09/110,716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 95
; TYPE: PRT
; ORGANISM: lipophilin C
US-09-110-716-31

Query Match      100.0%; Score 95; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 3, 2e-83;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 MKLLMVLMLAALLHCYADSCCKLEDMVEXTISDISIPEYKLLQEFIDSDAAARAG 60
DB      1 MKLLMVLMLAALLHCYADSCCKLEDMVEXTISDISIPEYKLLQEFIDSDAAARAG 60

OY      61 KFKOCFLNQSHTLKNFGIMHTYDSITCMKSN 95
DB      61 KFKOCFLNQSHTLKNFGIMHTYDSITCMKSN 95

RESULT 2
US-09-985-911-6
; Sequence 6, Application US/09985911
; Patent No. US20020151012A1
; GENERAL INFORMATION:
; APPLICANT: NI ET AL.
; TITLE OF INVENTION: HUMAN ENDOMETRIAL SPECIFIC STEROID-BINDING FACTOR I, II AND III
; FILE REFERENCE: PF257D3
; CURRENT APPLICATION NUMBER: US/09/985,911
```

```
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/583,169
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 09/263,810
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 08/821,451
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: 60/014,724
; PRIOR FILING DATE: 1996-03-21
; NUMBER OF SEQ. ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 95
; TYPE: PRT
; ORGANISM: human
US-09-985-911-6

Query Match      100.0%; Score 95; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 3.2e-83;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWVLMALALHGYADSGCKLEDMVEKTIISDIPIYKELQEFIDSDAAAEAMG 60
DB 1 MKLWVLMALALHGYADSGCKLEDMVEKTIISDIPIYKELQEFIDSDAAAEAMG 60
QY 61 KFKOCFLNOSHRTLNKFNGLMMHTVYDSIMCNMKS 95
DB 61 KFKOCFLNOSHRTLNKFNGLMMHTVYDSIMCNMKS 95

RESULT 3
US-09-110-716-13
; Sequence 13, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehner, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596.00
; CURRENT APPLICATION NUMBER: US/09/110,716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-110-716-13

Query Match      81.1%; Score 77; DB 10; Length 77;
Best Local Similarity 100.0%; Pred. No. 2.9e-66;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 DSGCKLEDMVEKTIISDIPIYKELQEFIDSDAAAEAMGKFKOCFLNOSHRTLNK 78
DB 1 DSGCKLEDMVEKTIISDIPIYKELQEFIDSDAAAEAMGKFKOCFLNOSHRTLNK 60
QY 79 LMHTVYDSIMCNMKS 95
DB 61 LMHTVYDSIMCNMKS 77

RESULT 4
US-09-110-716-40
; Sequence 40, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehner, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596.00
```

```
; CURRENT APPLICATION NUMBER: US/09/110,716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Lpnc
US-09-110-716-40

Query Match      68.4%; Score 65; DB 10; Length 76;
Best Local Similarity 100.0%; Pred. No. 6.4e-55;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 DSGCKLEDMVEKTIISDIPIYKELQEFIDSDAAAEAMGKFKOCFLNOSHRTLNK 78
DB 1 DSGCKLEDMVEKTIISDIPIYKELQEFIDSDAAAEAMGKFKOCFLNOSHRTLNK 60
QY 79 LMHT 83
DB 61 LMHT 65

RESULT 5
US-09-757-417-29
; Sequence 29, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-417-29

Query Match      12.6%; Score 12; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

RESULT 6
US-09-757-417-27
; Sequence 27, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 93
; TYPE: PRT
```

ORGANISM: Homo sapien
US-09-757-417-27

Query Match 12.6%: Score 12; DB 10; Length 93;
Best Local Similarity 100.0%: Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLMLVLMAL 12
Db 1 MKLMLVLMAL 12

RESULT 7
US-09-934-054-3
Sequence 3, Application US/09934054
Patent No. US20020107385A1
GENERAL INFORMATION:

APPLICANT: Akerblom, Ingrid E.
Hillman, Jennifer L.
Murry, Lynn E.
Goli, Surya K.

TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA

COUNTRY: USA
ZIP: 94025-6936

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/934,054
FILING DATE: 21-Aug-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,547

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PR-0077 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: <Unknown>

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-934-054-3

Query Match 12.6%: Score 12; DB 10; Length 93;
Best Local Similarity 100.0%: Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLMLVLMAL 12
Db 1 MKLMLVLMAL 12

RESULT 8
US-09-934-054-10
Sequence 10, Application US/09934054
Patent No. US20020107385A1
GENERAL INFORMATION:

APPLICANT: Akerblom, Ingrid E.
Hillman, Jennifer L.
Murry, Lynn E.
Goli, Surya K.

Hawkins, Phillip R.

TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA

COUNTRY: USA
ZIP: 94025-6936

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/934,054
FILING DATE: 21-Aug-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/747,547

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PR-0077 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: <Unknown>

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Query Match 12.6%: Score 12; DB 10; Length 93;
Best Local Similarity 100.0%: Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLMLVLMAL 12
Db 1 MKLMLVLMAL 12

RESULT 9
US-10-007-805-503
Sequence 503, Application US/10007805
Patent No. US20020150581A1
GENERAL INFORMATION:

APPLICANT: Jjiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

APPLICANT: Fanger, Gary R.

APPLICANT: Vedavick, Thomas S.

APPLICANT: McNeill, Patricia D.

APPLICANT: Durham, Margarita

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.470C10
CURRENT APPLICATION NUMBER: US/10/007,805
CURRENT FILING DATE: 2001-12-07
NUMBER OF SEQ ID NOS: 593

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 503
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-503
```

```
Query Match          12.6%; Score 12; DB 12; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 MKLMLVLMAL 12
| | | | | | | | | |
Db 1 MKLMLVLMAL 12
```

```
RESULT 10
US-09-757-417-47
; Sequence 47, Application US/09757417
; Patent No. US20020082216a1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY, DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479c1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-417-47
```

```
Query Match          12.6%; Score 12; DB 10; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 MKLMLVLMAL 12
| | | | | | | | | |
Db 40 MKLMLVLMAL 51
```

```
RESULT 11
US-10-007-805-495
; Sequence 495, Application US/10007805
; Patent No. US20020150581a1
; GENERAL INFORMATION:
```

```
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470c10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 495
; LENGTH: 410
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-007-805-495
```

```
Query Match          12.6%; Score 12; DB 12; Length 410;
Best Local Similarity 100.0%; Pred. No. 0.00039;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 MKLMLVLMAL 12
| | | | | | | | | |
Db 1 MKLMLVLMAL 12
```

```
RESULT 12
US-10-007-805-494
; Sequence 494, Application US/10007805
; Patent No. US20020150581a1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470c10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 494
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 403, 522, 615
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-007-805-494
```

```
Query Match          12.6%; Score 12; DB 12; Length 743;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 MKLMLVLMAL 12
| | | | | | | | | |
Db 1 MKLMLVLMAL 12
```

```
RESULT 13
US-10-007-805-493
; Sequence 493, Application US/10007805
; Patent No. US20020150581a1
; GENERAL INFORMATION:
```

```
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470c10
```

```

; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 403, 522, 615
; OTHER INFORMATION: Xaa - Any Amino Acid
US-10-007-805-493

```

```

Query Match          12.6%: Score 12; DB 12; Length 1095;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MKLIMVLMAL 12
        |||||
Db       1 MKLIMVLMAL 12

```

```

RESULT 14
US-09-757-417-4
; Sequence 4, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-757-417-4

```

```

Query Match          11.6%: Score 11; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      41 EYKELLOEFD 51
        |||||
Db       1 EYKELLOEFD 11

```

```

RESULT 15
US-09-757-417-33
; Sequence 33, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 20
; TYPE: PRT

```

```

; ORGANISM: Homo sapien
US-09-757-417-33

```

```

Query Match          11.6%: Score 11; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      41 EYKELLOEFD 51
        |||||
Db       1 EYKELLOEFD 11

```

```

Search completed: January 25, 2003, 13:39:19
Job time : 11 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:27:44 : Search time 18 Seconds
(without alignments)
155.288 Million cell updates/sec

Title: US-09-806-302a-2

Perfect score: 95
Sequence: 1 MKILMVLMLALLHCYADS.....NFCIMHTYYSIMCNKSN 95

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	95	100.0	95	3	US-08-821-451A-6	Sequence 6, Appl
2	95	100.0	95	4	US-09-263-810-6	Sequence 6, Appl
3	95	100.0	95	4	US-09-583-169-6	Sequence 6, Appl
4	12	12.6	93	1	US-08-455-896-2	Sequence 2, Appl
5	12	12.6	93	2	US-08-933-149-2	Sequence 2, Appl
6	12	12.6	93	2	US-09-082-343-2	Sequence 2, Appl
7	12	12.6	93	3	US-09-082-353-2	Sequence 2, Appl
8	12	12.6	93	4	US-09-215-818-5	Sequence 2, Appl
9	12	12.6	93	5	PCT-US96-08235-2	Sequence 2, Appl
10	7	7.4	528	3	US-08-403-852D-21	Sequence 21, Appl
11	7	7.4	528	3	US-08-510-646B-22	Sequence 22, Appl
12	7	7.4	528	4	US-09-231-818-21	Sequence 21, Appl
13	7	7.4	528	4	US-09-231-818-21	Sequence 21, Appl
14	6	6.3	11	1	US-07-699-468-2	Sequence 2, Appl
15	6	6.3	11	1	US-07-699-468-3	Sequence 3, Appl
16	6	6.3	11	1	US-07-699-468-4	Sequence 4, Appl
17	6	6.3	11	1	US-07-699-468-5	Sequence 5, Appl
18	6	6.3	11	1	US-07-699-468-6	Sequence 6, Appl
19	6	6.3	16	2	US-08-723-415B-9	Sequence 9, Appl
20	6	6.3	16	4	US-09-189-627A-9	Sequence 9, Appl
21	6	6.3	22	3	US-08-940-095-11	Sequence 11, Appl
22	6	6.3	22	3	US-08-940-095-17	Sequence 17, Appl
23	6	6.3	22	3	US-08-940-095-23	Sequence 23, Appl
24	6	6.3	22	3	US-08-940-095-36	Sequence 36, Appl
25	6	6.3	22	3	US-08-940-095-87	Sequence 87, Appl
26	6	6.3	22	3	US-08-940-093-11	Sequence 11, Appl
27	6	6.3	22	3	US-08-940-093-17	Sequence 17, Appl

28	6	6.3	22	3	US-08-940-093-23	Sequence 23, Appl
29	6	6.3	22	3	US-08-940-093-36	Sequence 36, Appl
30	6	6.3	22	3	US-08-940-093-87	Sequence 87, Appl
31	6	6.3	22	3	US-08-940-096-11	Sequence 11, Appl
32	6	6.3	22	3	US-08-940-096-17	Sequence 17, Appl
33	6	6.3	22	3	US-08-940-096-23	Sequence 23, Appl
34	6	6.3	22	3	US-08-940-096-36	Sequence 36, Appl
35	6	6.3	22	3	US-08-940-096-87	Sequence 87, Appl
36	6	6.3	22	4	US-09-465-719-11	Sequence 11, Appl
37	6	6.3	22	4	US-09-465-719-17	Sequence 17, Appl
38	6	6.3	22	4	US-09-465-719-23	Sequence 23, Appl
39	6	6.3	22	4	US-09-465-719-36	Sequence 36, Appl
40	6	6.3	22	4	US-09-465-719-87	Sequence 87, Appl
41	6	6.3	22	4	US-09-453-605-11	Sequence 11, Appl
42	6	6.3	22	4	US-09-453-605-17	Sequence 17, Appl
43	6	6.3	22	4	US-09-453-605-23	Sequence 23, Appl
44	6	6.3	22	4	US-09-453-605-36	Sequence 36, Appl
45	6	6.3	22	4	US-09-453-605-87	Sequence 87, Appl

ALIGNMENTS

RESULT 1
US-08-821-451A-6
Patent No. 6066724
GENERAL INFORMATION:
APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,451A
FILING DATE: March 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,724
FILING DATE: March 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (P257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-821-451A-6
Query Match 100.0%; Score 95; DB 3; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.8e-87;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Caps 0;
1 MKILMVLMLALLHCYADSCCKLEDMVEKTI NSDISIPYKELQEFIDSDAAEAMG 60

Db 1 M K L M V L M L A L L L H C Y A D S G C K L E D M V E K T I N S D I S I P E Y K E L L O E F I D S D A A E A M G 60
QY 61 K F K O C F L N O S H R T L K N G L M M H T Y Y D S I M C N M K S N 95
Db 61 K F K O C F L N O S H R T L K N G L M M H T Y Y D S I M C N M K S N 95

RESULT 2

US-09-263-810-6
Sequence 6, Application US/09263810

Patent No. 6174992

GENERAL INFORMATION:

APPLICANT: JIAN N1, Guo-Liang Yu and Reiner Gentz

TITLE OF INVENTION: Human Endometrial Specific Steroid-

TITLE OF INVENTION: Binding Factor I, II and III

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/263,810

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/821,451

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.

REGISTRATION NUMBER: 33,073

REFERENCE/DOCKET NUMBER: 325800-521 (PF257)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 95 AMINO ACIDS

TYPE: AMINO ACID

STRANDEDNESS:

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

US-09-263-810-6

Query Match 100.0%; Score 95; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.8e-87;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K L M V L M L A L L L H C Y A D S G C K L E D M V E K T I N S D I S I P E Y K E L L O E F I D S D A A E A M G 60
Db 1 M K L M V L M L A L L L H C Y A D S G C K L E D M V E K T I N S D I S I P E Y K E L L O E F I D S D A A E A M G 60
QY 61 K F K O C F L N O S H R T L K N G L M M H T Y Y D S I M C N M K S N 95
Db 61 K F K O C F L N O S H R T L K N G L M M H T Y Y D S I M C N M K S N 95

RESULT 3

US-09-583-169-6

Sequence 6, Application US/09583169

Patent No. 6338948

GENERAL INFORMATION:

APPLICANT: JIAN N1, Guo-Liang Yu and Reiner Gentz

TITLE OF INVENTION: Human Endometrial Specific Steroid-

TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/583,169

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/821,451

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.

REGISTRATION NUMBER: 33,073

REFERENCE/DOCKET NUMBER: 325800-521 (PF257)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 95 AMINO ACIDS

TYPE: AMINO ACID

STRANDEDNESS:

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

US-09-583-169-6

Query Match 100.0%; Score 95; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.8e-87;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K L M V L M L A L L L H C Y A D S G C K L E D M V E K T I N S D I S I P E Y K E L L O E F I D S D A A E A M G 60
Db 1 M K L M V L M L A L L L H C Y A D S G C K L E D M V E K T I N S D I S I P E Y K E L L O E F I D S D A A E A M G 60
QY 61 K F K O C F L N O S H R T L K N G L M M H T Y Y D S I M C N M K S N 95
Db 61 K F K O C F L N O S H R T L K N G L M M H T Y Y D S I M C N M K S N 95

RESULT 4

US-08-455-896-2

Sequence 2, Application US/08455896

Patent No. 5668267

GENERAL INFORMATION:

APPLICANT: WATSON, MARK A.

TITLE OF INVENTION: DNA SEQUENCE AND ENCODED

TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: ROGERS, HOWELL & HAFERKAMP

STREET: 7733 FORSYTH BOULEVARD, SUITE 1400

CITY: ST. LOUIS

STATE: MISSOURI

COUNTRY: USA

ZIP: 63105-1817

COMPUTER READABLE FORM:

COMPUTER: FLOPPY disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-455-896-2

```

```

Query Match      12.6%; Score 12; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

```

```

RESULT 5
US-08-933-149-2
Sequence 2, Application US/08933149
Patent No. 5922836
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,149
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: HENDERSON, MELDIE W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-933-149-2

```

```

Query Match      12.6%; Score 12; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

```

```

RESULT 6
US-09-082-343-2
Sequence 2, Application US/09082343
Patent No. 5968754
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,343
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/455,896
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-082-343-2

Query Match      12.6%; Score 12; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

RESULT 7
US-09-082-253-2
Sequence 2, Application US/09082253
Patent No. 6004756
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
MAMMARY-SPECIFIC BREAST CANCER PROTEIN

```

NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,253
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/455,896
FILING DATE: 05/31/1995
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
US-09-082-253-2

Query Match 12.6%; Score 12; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

RESULT 8
US-09-215-818-5
Sequence 5, Application US/09/215818A
Patent No. 6379671
GENERAL INFORMATION:
APPLICANT: Colpitts, Tracey
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
TITLE OF INVENTION: DETECTING DISEASES OF THE BREAST
FILE REFERENCE: 5972.US.P2
CURRENT APPLICATION NUMBER: US/09/215,818A
CURRENT FILING DATE: 1998-12-18
EARLIER APPLICATION NUMBER: 08/912,276
EARLIER FILING DATE: 1997-08-17
EARLIER APPLICATION NUMBER: 08/697,105
EARLIER FILING DATE: 1996-08-19
EARLIER APPLICATION NUMBER: 08/2912,149
EARLIER FILING DATE: 1997-08-15
EARLIER APPLICATION NUMBER: 08/697,106
EARLIER FILING DATE: 1996-08-19
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 93
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-215-818-5

Query Match 12.6%; Score 12; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

RESULT 9
PCT-US96-08235-2
Sequence 2, Application PCT/US9608235
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08235
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 964796
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
PCT-US96-08235-2

Query Match 12.6%; Score 12; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLWVLMAL 12
DB 1 MKLWVLMAL 12

RESULT 10
US-08-403-852D-21
Sequence 21, Application US/08403852D
Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibault, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent

```

1      APPLICANT: De Crecy-Lagard, Valerie
2      TITLE OF INVENTION: Polypeptides Involved In The
3      TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
4      TITLE OF INVENTION: Coding For These Polypeptides And Their Use
5      NUMBER OF SEQUENCES: 43
6      CORRESPONDENCE ADDRESS:
7      ADDRESSEE: Flinegan, Henderson, Farbow, Garrett & Dunner
8      STREET: 1300 I Street, N.W., Suite 700
9      CITY: Washington
10     STATE: D.C.
11     COUNTRY: USA
12     ZIP: 20005-3315
13     COMPUTER READABLE FORM:
14     MEDIUM TYPE: Floppy disk
15     COMPUTER: IBM PC compatible
16     OPERATING SYSTEM: PC-DOS/MS-DOS
17     SOFTWARE: Patent In Release #1.0, Version #1.30
18     CURRENT APPLICATION DATA:
19     APPLICATION NUMBER: US/08/403,852D
20     FILING DATE: 10-MAY-1995
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER: PCT/FR 93/00923
23     FILING DATE: 25-SEP-1993
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER: FR 92/11441
26     FILING DATE: 25-SEP-1992
27     ATTORNEY/AGENT INFORMATION:
28     NAME: Meyers, Kenneth J.
29     REGISTRATION NUMBER: 25,146
30     REFERENCE/DOCKET NUMBER: 03806.0054-00000
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: (202) 408-4000
33     TELEFAX: (202) 408-4400
34     INFORMATION FOR SEQ ID NO: 21:
35     SEQUENCE CHARACTERISTICS:
36     LENGTH: 528 amino acids
37     TYPE: amino acid
38     STRANDEDNESS: single
39     TOPOLOGY: linear
40     US-08-403-852D-21
41
42     Query Match 7.4%; Score 7; DB 2; Length 528;
43     Best Local Similarity 100.0%; Pred. No. 33;
44     Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
45
46     QY 9 LAALLLH 15
47         |||||
48         501 LAALLLH 507
49
50 RESULT 11
51 US-08-510-646B-22
52 : Sequence 22, Application US/08510646B
53 : Patent No.6077699
54 : GENERAL INFORMATION:
55 : APPLICANT: Blanc, Veronique
56 : APPLICANT: Branche, Francis
57 : APPLICANT: Crouzet, Joel
58 : APPLICANT: Jacques, Nathalie
59 : APPLICANT: Lacroix, Patricia
60 : APPLICANT: Thibaut, Denis
61 : APPLICANT: Zagorec, Monique
62 : APPLICANT: Debussche, Laurent
63 : APPLICANT: De Crecy-Lagard, Valerie
64 : TITLE OF INVENTION: Polypeptides Involved In The
65 : TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
66 : TITLE OF INVENTION: Coding For These Polypeptides And Their Use
67 : NUMBER OF SEQUENCES: 45
68 : CORRESPONDENCE ADDRESS:
69 : ADDRESSEE: Flinegan, Henderson, Farbow, Garrett & Dunner
70 : STREET: 1300 I Street, N.W., Suite 700
71 : CITY: Washington
72 : STATE: D.C.
73

```

```

1 COUNTRY: USA
2 ZIP: 20005-3315
3
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: Floppy disk
6 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8 SOFTWARE: Patent In Release #1.0, Version #1.30
9
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/510,646B
12 FILING DATE: 03-AUG-1995
13 CLASSIFICATION: 435
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: US 08/403,852
16 FILING DATE: 10-MAY-1995
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: PCT/FR 93/00923
19 FILING DATE: 25-SEP-1993
20 APPLICATION DATA:
21 APPLICATION NUMBER: FR 92/11441
22 FILING DATE: 25-SEP-1992
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Meyers, Kenneth J.
25 REGISTRATION NUMBER: 25,146
26 REFERENCE/DOCKET NUMBER: 03806,0054-01000
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (202) 408-4000
29 TELEFAX: (202) 408-4400
30 INFORMATION FOR SEQ ID NO: 22:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 528 amino acids
33 TYPE: amino acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36
37 US-08-510-646B-22
38
39 Query Match 7.4%; Score 7; DB 3; Length 528;
40 Best Local Similarity 100.0%; Pred. No. 33;
41 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
42
43 QY 9 LALLLH 15
44 |||||
45 Db 501 LALLLH 507
46
47 RESULT 12
48 US-09-231-818-21
49 Sequence 21, Application US/09231818
50 Patent No. 6171846
51
52 GENERAL INFORMATION:
53 APPLICANT: Blanc, Veronique
54 APPLICANT: Blanche, Francis
55 APPLICANT: Crouzet, Joel
56 APPLICANT: Jacques, Nathalie
57 APPLICANT: Lacroix, Patricia
58 APPLICANT: Thibaut, Denis
59 APPLICANT: Zagorec, Monique
60 APPLICANT: Debussche, Laurent
61 APPLICANT: De Crecy-Lagard, Valerie
62 TITLE OF INVENTION: Polypeptides Involved In The
63 TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
64 TITLE OF INVENTION: Coding For These Polypeptides And Their Use
65 NUMBER OF SEQUENCES: 43
66 CORRESPONDENCE ADDRESS:
67 ADDRESSEE: Finnegan, Henderson, Farbow, Garrett & Dunner
68 STREET: 1300 I Street, N.W., Suite 700
69 CITY: Washington
70 STATE: D.C.
71 COUNTRY: USA
72 ZIP: 20005-3315
73
74 COMPUTER READABLE FORM:
75 MEDIUM TYPE: Floppy disk
76 COMPUTER: IBM PC compatible
77 OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231,818
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-231-818-21

Query Match 7.4%; Score 7; DB 4; Length 528;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LAALLH 15
|||||
Db 501 LAALLH 507

RESULT 13
US-07-699-468-2
Sequence 2, Application US/07699468
Patent No. 5350574
GENERAL INFORMATION:
APPLICANT: Erlanger, Bernard F
APPLICANT: Cicalano, William L
APPLICANT: Cicalano, Nicholas A
TITLE OF INVENTION: NOVEL DERIVATIVES OF CYCLOSPORINE A,
TITLE OF INVENTION: ANTIBODIES DIRECTED THERETO AND USES
TITLE OF INVENTION: THEROF
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/699,468
FILING DATE: 19910513
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 280009
FILING DATE: 05-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 33654-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
TOPOLOGY: circular
MOLECULE TYPE: peptide
HYPOTHEICAL: N
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note="Calpha-methyl,trans-2-butadiene"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /note="Cbeta-methyl"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note="Methyl"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note="Methyl"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /note="Methyl"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 8
OTHER INFORMATION: /note="Calpha-R"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /note="Methyl"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 10
OTHER INFORMATION: /note="Methyl"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 11
OTHER INFORMATION: /note="Methyl"
US-07-699-468-2

Query Match 6.3%; Score 6; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LAALL 14
|||||
Db 6 LAALL 11

RESULT 14
US-07-699-468-3
Sequence 3, Application US/07699468
Patent No. 5350574
GENERAL INFORMATION:
APPLICANT: Erlanger, Bernard F
APPLICANT: Cicalano, William L
APPLICANT: Cicalano, Nicholas A
TITLE OF INVENTION: NOVEL DERIVATIVES OF CYCLOSPORINE A,
TITLE OF INVENTION: ANTIBODIES DIRECTED THERETO AND USES
TITLE OF INVENTION: THEROF
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York

STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/699,468
FILING DATE: 19910513
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 280009
FILING DATE: 05-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 33654-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
TOPOLOGY: circular
MOLECULE TYPE: peptide
HYPOTHETICAL: N
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note=
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note="MeGly"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note="MeLeu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /note="MeLeu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 8
OTHER INFORMATION: /note="Calpha-R"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /note="MeLeu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 10
OTHER INFORMATION: /note="MeLeu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 11
OTHER INFORMATION: /note="MeLeu"
US-07-699-468-3

Query Match 6.38; Score 6; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 LAALLL 14
| | | | |
Db 6 LAALLL 11

RESULT 15
US-07-699-468-4
Sequence 4, Application US/07699468
Patent No. 5350574
GENERAL INFORMATION:
APPLICANT: Erlanger, Bernard F
APPLICANT: Cleveland, William L
APPLICANT: Cacalano, Nicholas A
TITLE OF INVENTION: NOVEL DERIVATIVES OF CYCLOSPORINE A,
ANTIBODIES DIRECTED THERETO AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/699,468
FILING DATE: 19910513
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 280009
FILING DATE: 05-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 33654-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
TOPOLOGY: circular
MOLECULE TYPE: peptide
HYPOTHETICAL: N
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note=
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note="MeGly"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note="MeLeu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /note="MeLeu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 8
OTHER INFORMATION: /note="Calpha-R"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /note="MeLeu"
FEATURE:

NAME/KEY: Modified-site
LOCATION: 10
OTHER INFORMATION: /note= "Meleu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 11
OTHER INFORMATION: /note= "Meleu"
US-07-699-468-4

Query Match 6.38; Score 6; DB 1; Length 11;
Best Local Similarity 100.08; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LAALLL 14
| | | | |
Db 6 LAALLL 11

Search completed: January 25, 2003, 13:35:50
Job time : 19 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

```
Run on:      January 25, 2003, 13:35:25 ; Search time 36 Seconds
              (without alignments)
              351.634 Million cell updates/sec
```

Title: US-09-806-302A-2

Sequence: 1 MKLIMVLMIALLLHCYADS.....NEGIMMHTVYDSIMCNMKS 95

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

word size : 0

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002:*

1:	/SID2/gcgdata/genseq/genseqp-emb1/AA1980.DAT
2:	/SID2/gcgdata/genseq/genseqp-emb1/AA1981.DAT
3:	/SID2/gcgdata/genseq/genseqp-emb1/AA1982.DAT
4:	/SID2/gcgdata/genseq/genseqp-emb1/AA1983.DAT
5:	/SID2/gcgdata/genseq/genseqp-emb1/AA1984.DAT
6:	/SID2/gcgdata/genseq/genseqp-emb1/AA1985.DAT
7:	/SID2/gcgdata/genseq/genseqp-emb1/AA1986.DAT
8:	/SID2/gcgdata/genseq/genseqp-emb1/AA1987.DAT
9:	/SID2/gcgdata/genseq/genseqp-emb1/AA1988.DAT
10:	/SID2/gcgdata/genseq/genseqp-emb1/AA1989.DAT
11:	/SID2/gcgdata/genseq/genseqp-emb1/AA1990.DAT
12:	/SID2/gcgdata/genseq/genseqp-emb1/AA1991.DAT
13:	/SID2/gcgdata/genseq/genseqp-emb1/AA1992.DAT
14:	/SID2/gcgdata/genseq/genseqp-emb1/AA1993.DAT
15:	/SID2/gcgdata/genseq/genseqp-emb1/AA1994.DAT
16:	/SID2/gcgdata/genseq/genseqp-emb1/AA1995.DAT
17:	/SID2/gcgdata/genseq/genseqp-emb1/AA1996.DAT
18:	/SID2/gcgdata/genseq/genseqp-emb1/AA1997.DAT
19:	/SID2/gcgdata/genseq/genseqp-emb1/AA1998.DAT
20:	/SID2/gcgdata/genseq/genseqp-emb1/AA1999.DAT
21:	/SID2/gcgdata/genseq/genseqp-emb1/AA2000.DAT
22:	/SID2/gcgdata/genseq/genseqp-emb1/AA2001.DAT
23:	/SID2/gcgdata/genseq/genseqp-emb1/AA2002.DAT

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	95	100.0	95	18	AAW35804	Human endometrial
2	95	100.0	95	20	AAW02590	A human mammsglobl
3	95	100.0	95	21	AAB03769	Human endometrial
4	95	100.0	95	21	AAI99226	Human endometrial
5	95	100.0	95	21	AAI92237	Human endometrial
6	95	100.0	95	21	AAW6594	Mammoglobin homold
7	95	100.0	95	22	AAB31682	Human 5' EST relat
8	95	100.0	95	23	AAO20555	An endometrial spe
9	95	100.0	95	23	ABB09635	Protein of human I
10	95	100.0	108	20	AAV60038	Human endometrial

11	73	16.8	74	19	AA651649	Non-ocular diseases
12	12	12.6	13	22	AA651129	Human mammaglobin
13	12	12.6	90	22	AAE07528	Human mammaglobin
14	12	12.6	90	22	AAE07537	Human mammaglobin
15	12	12.6	93	18	AAW10119	Mammary-specific s
16	12	12.6	93	19	AAW59777	Amino acid sequenc
17	12	12.6	93	19	AAW48432	Mammaglobin protei
18	12	12.6	93	20	AAV01718	Mammaglobin, a mam
19	12	12.6	93	21	AAAB13786	Human mammaglobin
20	12	12.6	93	21	AAW84522	Amino acid sequenc
21	12	12.6	93	22	AAE07517	Human mammaglobin
22	12	12.6	93	22	AAE07529	Human mammaglobin
23	12	12.6	93	22	AAE07530	Human mammaglobin
24	12	12.6	93	22	AAE07531	Human mammaglobin
25	12	12.6	93	22	AAE07532	Human mammaglobin
26	12	12.6	93	22	AAE07533	Human mammaglobin
27	12	12.6	93	22	AAE07535	Human mammaglobin
28	12	12.6	93	22	AAE07536	Human mammaglobin
29	12	12.6	93	22	AAE07537	Human mammaglobin
30	12	12.6	410	22	AAE07538	Human mammaglobin
31	12	12.6	743	22	AAU33359	Human breast cancer
32	12	12.6	1095	22	AAU33358	Human breast cancer
33	12	11.6	20	22	AAU33357	Human breast cancer
34	11	11.6	20	22	AAU33356	Human mammaglobin
35	11	11.6	20	22	AAU33355	Human mammaglobin
36	11	11.6	74	21	AAW48433	Mammaglobin synthet
37	11	11.6	93	22	AAE07534	Amino acid sequenc
38	11	11.6	220	22	AAE07535	Human mammaglobin
39	10	10.5	10	22	AAO22141	Ra12-mammaglobin f
40	10	10.5	20	22	AAE07536	Human mammaglobin
41	9	9.5	20	22	AAE07537	Human mammaglobin
42	9	9.5	9	22	AAE07538	Human mammaglobin
43	9	9.5	9	22	AAE07539	Human mammaglobin
44	9	9.5	9	22	AAE07540	Human mammaglobin
45	8	8.4	20	19	AAW48434	Mammaglobin synthet

ALIGNMENTS

XX	AAW35804	Standard; Protein; 95 AA.
XX	AAW35804;	
XX	27-MAR-1998	(First entry)
XX	Human endometrial specific steroid-binding factor III.	
KW	Endometrial specific steroid-binding factor III; ESF III; human;	
KW	Clara cell secretory protein; endometrium;	
KM	phospholipase A2 inhibitor; polychlorinated biphenyl; antiagregant;	
KM	inflammation; asthma; rhinitis; cystic fibrosis; airway disease;	
KM	neoplasia; atopy; therapy; diagnosis.	
OS	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
FH	Peptide	1..21
FT	Protein	/label= Sig_peptide
FT		22..95
FT		/label= Mat_protein
XX		
PN	W09734997-A1.	
XX		
XX	25-SEP-1997.	
PD		
XX		
PF	21-MAR-1996;	96WO-US03857.
XX		
PR	21-MAR-1996;	96WO-US03857.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	

XX Genz RL, Ni J, Yu G;
 XX
 DR WPI: 1997-480206/44.
 DR N-PSDB: AAT94832.
 XX
 PT Human endometrial specific steroid-binding factor I, II and III -
 PT used to treat inflammation, asthma, rhinitis, cystic fibrosis,
 PT allergy disease, neoplasia, atopy etc.
 XX
 PS Claim 19; Page 65; 92pp; English.
 XX
 CC This sequence comprises human endometrial specific steroid binding
 CC factor III (ESF III), a protein that inhibits phospholipase A2
 CC activity, binds to polychlorinated biphenyl compounds, reduces
 CC foreign protein antigenicity, inhibits monocyte and neutrophil
 CC chemotaxis and phagocytosis, inhibits platelet aggregation,
 CC regulates eicosanoid levels in the human uterus and controls the
 CC growth of endometrial cells. The amino acid sequence was deduced
 CC from a cDNA clone (see AAT94832) derived from a human endometrial
 CC tumour. ESF I (see AAM35802) and ESF II (see AAM35803) are also
 CC claimed. Human ESF III has about 36% identity with rat prostatic
 CC steroid-binding protein. Recombinant ESF I, II and III can be
 CC expressed in host cells for use in claimed methods (a) for treating
 CC a patient in need of ESF I, II or III (including expression of the
 CC polypeptide in vivo) and (b) for identifying compounds which bind
 CC to and inhibit activation of the ESF polypeptide. hESF I, II and
 CC III may be used to treat inflammation, asthma, rhinitis, cystic
 CC fibrosis, allergy disease, neoplasia and atopy.
 XX
 SQ Sequence 95 AA;

Query Match 100.0%; Score 95; DB 18; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1.5e-89;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLLVLMALALLHCAVDSGCKLLEDVVEKTSIDISPEYKELLOEFTDSPAAPAWG 60
 DB 1 MKLLVLMALALLHCAVDSGCKLLEDVVEKTSIDISPEYKELLOEFTDSPAAPAWG 60
 OY 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIWCNMSN 95
 DB 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIWCNMSN 95

RESULT 2
 AAY02590
 ID AAY02590 standard; Protein; 95 AA.
 XX
 AC AAY02590;
 XX

DT 26-JUL-1999 (first entry)
 XX
 DE A human mammary globin homologue (HMH).

XX Human mammary globin homologue; HMH; antagonist; neoplastic disorder;
 KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
 KW teratocarcinoma; endometriosis.
 XX

OS Homo sapiens.
 XX
 PN WO919487-A1.
 XX
 PD 22-APR-1999.
 XX
 PF 14-OCT-1998; 98WO-US21729.
 XX
 PR 16-OCT-1997; 97US-0951750.
 XX

PA (INCY-) INCYTE PHARM INC.
 XX
 PI Hillman JL, Murry LE, Shah P;
 XX

DR WPI: 1999-302531/25.
 DR N-PSDB: AAX36138.
 XX
 PT New human mammary globin homolog (HMH), useful for diagnosing, treating
 PT or preventing disorders associated with expression of HMH
 XX
 PS Claim 1; Fig 1A-B; 63pp; English.
 XX

CC The present sequence represents a human mammary globin homologue (HMH).
 CC Antagonists of the HMH polypeptide can be used to treat neoplastic
 CC disorders including adenocarcinoma, leukemia, lymphoma, melanoma,
 CC myeloma, sarcoma and teratocarcinoma. A vector expressing the
 CC complement of the polynucleotide encoding HMH may be administered
 CC to a subject to treat or prevent neoplastic disorders or endometriosis.
 CC Antibodies which bind HMH may also be used in the diagnosis of
 CC conditions or diseases characterized by expression of HMH, or in assays
 CC to monitor patients being treated with HMH, agonists, antagonists or
 CC inhibitors. Polynucleotides encoding HMH may also be used diagnostically
 CC to detect and quantitate gene expression in biopsied tissues. With
 CC respect to cancer a relatively high amount of transcript may indicate a
 CC predisposition for the development of disease. The nucleic acid sequences
 CC which encode HMH may also be used to generate hybridization probes useful
 CC for mapping the naturally occurring genomic sequence. HMH, and its
 CC fragments/variants can be used for screening libraries of compounds in
 CC drug screening techniques.
 XX

SQ Sequence 95 AA;
 Query Match 100.0%; Score 95; DB 20; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1.5e-89;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLLVLMALALLHCAVDSGCKLLEDVVEKTSIDISPEYKELLOEFTDSPAAPAWG 60
 DB 1 MKLLVLMALALLHCAVDSGCKLLEDVVEKTSIDISPEYKELLOEFTDSPAAPAWG 60
 OY 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIWCNMSN 95
 DB 61 KFKOCFLNOSHRTLNKNGMLMHTVYDSIWCNMSN 95

RESULT 3
 AAB03769
 ID AAB03769 standard; Protein; 95 AA.
 XX
 AC AAB03769;
 XX

DT 06-OCT-2000 (first entry)
 XX

DE Human endometrial specific steroid-binding factor III protein sequence.

XX Endometrial specific steroid-binding factor; human; hESF; inflammation;
 KW asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; atopy;
 KW eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.
 XX

OS Homo sapiens.
 XX
 PN US6066724-A.
 XX

PD 23-MAY-2000.
 XX
 PF 21-MAR-1997; 97US-0821451.
 XX

PR 21-MAR-1996; 96US-0014724.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Yu G, Genz R, Ni J;
 XX
 DR WPI: 2000-375600/32.
 DR N-PSDB: AAS59730.
 XX

PT Novel gene encoding human endometrial specific steroid-binding factor

PT I, II and III which is useful for treating asthma, rhinitis, cystic
PT fibrosis, airway disease and neoplasia
PS
PS Claim 1: Fig 3: 36pp: English.

CC This invention relates to nucleic acid molecules encoding portions of the
CC human endometrial specific steroid-binding factors I, II, and III. Also
CC included in the invention are hsf I, II, and III polypeptide sequences.
CC The nucleotide sequence exhibit antiasthmatic, antiinflammatory,
CC anti-allergic, and cytostatic properties. The polynucleotides are used in
CC gene therapy to express hsf I, II and III polypeptides in vivo to treat
CC and/or prevent inflammation, asthma, rhinitis, cystic fibrosis, air way
CC disease, neoplasia and atopy. The polynucleotides are also used to
CC inhibit phospholipase A2 activity, bind polychlorinated biphenyls, reduce
CC foreign protein antigenicity, inhibit monocyte and neutrophil chemotaxis
CC and phagocytosis, inhibit platelet aggregation, regulate eicosanoid
CC levels in the human uterus and control the growth of endometrial cells.
CC The polynucleotides are also useful for detecting complementary
CC polynucleotides as a diagnostic reagent. The hsf I, II and III
CC polynucleotides are used to detect complementary polynucleotides such as
CC a diagnostic reagent. Detection of a mutated form of hsf I, II and III
CC associated with a dysfunction will provide a diagnostic tool that can
CC define diagnosis of a disease or susceptibility to a disease which
CC results from under-expression, over-expression or altered expression of
CC hsf I, II and III e.g. a susceptibility to inherited asthma and
CC endometrial cancer. They are also useful for chromosome identification.
CC The present sequence represents a hsf III protein sequence identified in
CC the invention.

SO Sequence 95 AA:

Query Match 100.0%; Score 95; DB 21; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.5e-89;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLIMVLMALLLHCYADSGCKLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60
DB 1 MKLIMVLMALLLHCYADSGCKLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60

QY 61 KFKOCFLNOSHRTLKNFGLMMHTVYDSIMCNMKN 95
DB 61 KFKOCFLNOSHRTLKNFGLMMHTVYDSIMCNMKN 95

RESULT 4
AAV92226

ID AAV92226 standard; Protein: 95 AA.

AC AAV92226;

DT 10-AUG-2000 (first entry)

DE Human endometrial specific steroid binding factor III.

KW Endometrial specific steroid binding factor; ESBPIII; diagnosis;
KM gynaecological cancer; uterine; breast; endometrial; ovarian;
KM antibody; gene therapy; cytostatic.

OS Homo sapiens.

PN WO200020044-A1.

PD 13-APR-2000.

PF 30-SEP-1999; 99WO-US22753.

PR 02-OCT-1998; 98US-0102743.

PA (DIAD-) DIADEXUS LLC.

PI Macina RA;

DR WPI: 2000-303649/26.

DR N-PSDB; AAA09104.

XX Diagnosing, staging and monitoring gynecological cancer comprising
PT using an elevated level of ESBPIII in a patient as an indicator of
PT cancer

PS Claim 6; Page 28-29; 32pp: English.

CC The levels of human endometrial specific steroid binding factor
CC (ESBPIII) can be measured and compared to control levels and used to
CC diagnose the presence of a gynaecological (uterine, breast, endometrial,
CC or ovarian) cancer in a patient. ESBPIII levels can also be used to
CC diagnose metastasis, to stage or monitor gynaecological cancer.
CC Antibodies specific for ESBPIII can be used to treat gynecological
CC cancers.

SO Sequence 95 AA:

Query Match 100.0%; Score 95; DB 21; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.5e-89;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLIMVLMALLLHCYADSGCKLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60
DB 1 MKLIMVLMALLLHCYADSGCKLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60

QY 61 KFKOCFLNOSHRTLKNFGLMMHTVYDSIMCNMKN 95
DB 61 KFKOCFLNOSHRTLKNFGLMMHTVYDSIMCNMKN 95

RESULT 5
AAV92237

ID AAV92237 standard; Protein: 95 AA.

AC AAV92237;

DT 10-AUG-2000 (first entry)

DE Mammoglobin homologue from clone Mamm-X.

KW Clone Mamm-X; mammoglobin; breast cancer; cytostatic; anti-HIV;
KM immunosuppressive; anti-allergic; anti-infective; anti-inflammatory;
KM anti-arthritic; anti-arteriosclerotic; vasotropic; neuroprotective;
KM nootropic; dermatological; tranquilizer; vulnerary.

OS Homo sapiens.

PN WO200020447-A2.

PD 13-APR-2000.

PF 06-OCT-1999; 99WO-US23294.

PR 06-OCT-1998; 98US-0103195.

PR 05-OCT-1999; 99US-0103195.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA;

DR WPI: 2000-303741/26.

DR N-PSDB; AAA09118.

PT Nucleic acids encoding polypeptides with syncollin-like, claudin-like or
PT cytokine-like activity, useful for treating diseases including cancer,
PT Alzheimer's and atherosclerosis

PS Claim 23; Fig 6; 118pp: English.

CC Clone Mamm-X encodes a polypeptide that is 100 percent identical to human
CC Mammoglobin B precursor, a potential marker of breast cancer nodal
CC metastasis. The sequences are useful for treatment of diseases such as

CC cancer, immune disorders, autoimmune disease, acquired immune deficiency
 CC syndrome (AIDS), transplant rejection, allergy, infection by a
 CC pathological agent or organism, inflammatory disorders, arthritis, a
 CC hematopoietic disorder, a skin disorder, atherosclerosis, restenosis, a
 CC neurological disease, Alzheimer's disease, trauma, spinal cord injury
 CC and skeletal disorders.

CC Sequence 95 AA:

Query Match 100.0%; Score 95; DB 21; Length 95;

Best Local Similarity 100.0%; Pred. No. 1.5e-89;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLAVLMLAALLHGYADSGCKLLEDWEKTIINSISIPYKELIQEFDSDAAAEAMG 60

DB 1 MKLLAVLMLAALLHGYADSGCKLLEDWEKTIINSISIPYKELIQEFDSDAAAEAMG 60

QY 61 KFKQCFLNOSHRTLNKNGGLMHTVYDSIMCNMKS 95

DB 61 KFKQCFLNOSHRTLNKNGGLMHTVYDSIMCNMKS 95

RESULT 6
 AAT65394

ID AAY65394 standard; Protein: 95 AA.

AC AAY65394;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1555.

KM Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;

KM gene therapy; chromosome mapping; upstream regulatory sequence;

KM forensic; location; development; protein synthesis; stability;

KM regulation; identification.

OS Homo sapiens.

PN WO993051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

PI Dumas MLine Edwards J, Duclert A, Giordano J;

DR WPI: 2000-038446/03.

DR N-PSDB: AA243008.

PT Novel secreted protein 5' expressed sequence tag sequences used in

PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS Claim 3: Page 818; 837pp; English.

XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)

CC sequences, corresponding to human secreted proteins. AAY64651 to

CC AAY6438 represent the EST-related proteins corresponding to AA242265 to

CC AA243052. The 5' ESTs can be used for producing secreted human gene

CC products. They can be used to identify and isolate 5' untranslated

CC regions (UTRs) and upstream regulatory regions which control the

CC location, development stage, rate, and quantity of protein synthesis, as

CC well as stability of mRNA. The ESTs are also useful as probes for

CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can

CC also be used in forensic procedures to identify individuals, or in

CC diagnostic procedures to identify individuals having genetic diseases

CC resulting from abnormal gene expression. The products may also be used in

CC gene therapy protocols. The nucleic acids encoding signal peptides can be

CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.

CC Sequence 95 AA:

Query Match 100.0%; Score 95; DB 21; Length 95;

Best Local Similarity 100.0%; Pred. No. 1.5e-89;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLAVLMLAALLHGYADSGCKLLEDWEKTIINSISIPYKELIQEFDSDAAAEAMG 60

DB 1 MKLLAVLMLAALLHGYADSGCKLLEDWEKTIINSISIPYKELIQEFDSDAAAEAMG 60

QY 61 KFKQCFLNOSHRTLNKNGGLMHTVYDSIMCNMKS 95

DB 61 KFKQCFLNOSHRTLNKNGGLMHTVYDSIMCNMKS 95

RESULT 7
 AAB31682

ID AAB31682 standard; Protein: 95 AA.

AC AAB31682;

DT 30-APR-2001 (first entry)

DE An endometrial specific steroid binding factor III.

KM Human; endometrial specific steroid binding factor; hESF; hESFI; hESFIT;

KM hESFII; inflammation; asthma; rhinitis; cystic fibrosis; airway disease;

KM neoplasia; atopy; phospholipase A2; polychlorinated biphenyl; chemotaxis;

KM phagocytosis; platelet aggregation; eicosanoid; endometrial cell.

OS Homo sapiens.

FN Key Location/Qualifiers

FT Peptide 1..21

FT /note="signal peptide"

PN US6174992-B1.

PD 16-JAN-2001.

PF 08-MAR-1999; 99US-0263810.

PR 21-MAR-1996; 96US-0014724.

PR 21-MAR-1997; 97US-0821451.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ni J, Yu G, Gentz R;

DR WPI: 2001-158477/16.

DR N-PSDB: AAF25214.

PT New human endometrial specific steroid binding factors, useful for

PT treating and preventing inflammation, asthma, rhinitis, cystic

PT fibrosis, airway disease, neoplasia and atopy

PS Claim 1: Fig 3; 36pp; English.

XX The present sequence represents a human endometrial specific steroid

CC binding factor (hESF). The specification describes hESFI, hESFII, and

CC hESFIII. hESFI, II and III polypeptides, and polynucleotides encoding

CC them are useful for treating and preventing inflammation, asthma,

CC rhinitis, cystic fibrosis, airway disease, neoplasia and atopy,

CC inhibiting phospholipase A2 activity, binding polychlorinated

CC biphenyls, reducing foreign protein antigenicity, inhibiting monocyte

CC abd neutrophil chemotaxis and phagocytosis, inhibiting platelet
 CC aggregation, regulating eicosanoid levels in the human uterus, and for
 CC controlling the growth of endometrial cells. hESF polypeptides and
 CC nucleotides are also useful for research, biological, clinical or
 CC therapeutic purposes.

XX Sequence 95 AA:

Query Match 100.0%; Score 95; DB 22; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1.5e-89;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLIMVLMALALLHCYADSGCKLLEDMEVKTIINSISPEYKELLOEFIDSDAAAEAMG 60
 DB 1 MKLIMVLMALALLHCYADSGCKLLEDMEVKTIINSISPEYKELLOEFIDSDAAAEAMG 60

OY 61 KFKOCFLNOSHRTLNKNGLMHHTVYDSIMCNMKS 95
 DB 61 KFKOCFLNOSHRTLNKNGLMHHTVYDSIMCNMKS 95

RESULT 8

AAO20555
 ID AAO20555 standard; Protein; 95 AA.

XX AAO20555;

XX 27-JUN-2002 (first entry)

XX Protein of human lipophilin C.

XX Immunogenic epitope; hormonally regulated organ; malignant tumour;

XX Lipophilin; human.

XX Homo sapiens.

XX US2002034739-A1.

XX 21-MAR-2002.

XX 07-JUL-1998; 980S-0110716.

XX 07-JUL-1998; 980S-0110716.

XX (LEHR/) LEHRER R I.

XX (ZHAO/) ZHAO C.

XX (GLAS/) GLASGOW B J.

XX Lehrer RI, Zhao C, Glasgow BJ;

XX WPI: 2002-338922/37.

XX N-PSDB: AAK99492.

XX Peptides having the sequence of human lipophilin A, B and C are

XX PT associated with carcinomas of hormonally regulated organs and are

XX PT useful in the diagnosis and prognosis of various cancers -

XX Claim 7: Fig 6; 22pp; English.

XX The invention relates to a peptide comprising the amino acid sequences of

XX CC human lipophilin A, B, or C or its allelic variant or fragment comprising

XX CC at least one immunogenic epitope, which is purified and isolated, and may

XX CC have the N-terminal acylated and/or C-terminal amidated or be a fusion

XX CC protein. Molecules of the invention are used in the diagnosis and

XX CC prognosis of malignant tumours associated with hormonally regulated

XX CC organs such as uterus, ovary, prostate, testis, breast, kidney and

XX CC thymus. This sequence represents the human lipophilin C protein of the

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLIMVLMALALLHCYADSGCKLLEDMEVKTIINSISPEYKELLOEFIDSDAAAEAMG 60
 DB 1 MKLIMVLMALALLHCYADSGCKLLEDMEVKTIINSISPEYKELLOEFIDSDAAAEAMG 60

OY 61 KFKOCFLNOSHRTLNKNGLMHHTVYDSIMCNMKS 95
 DB 61 KFKOCFLNOSHRTLNKNGLMHHTVYDSIMCNMKS 95

RESULT 9

ABBO9635
 ID ABBO9635 standard; Protein; 95 AA.

XX ABBO9635;

XX 29-MAY-2002 (first entry)

XX Human endometrial specific steroid-binding factor (hESF) III.

XX Human endometrial specific steroid-binding factor; ESF;

XX KW prostatic steroid-binding protein; hESF I; hESF II; hESF III; asthma.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT Peptide 1..21

XX FT Protein /note="signal peptide"

XX FT 22..95

XX FT /note="mature protein"

XX US6338948-B1.

XX 15-JAN-2002.

XX 30-MAY-2000; 2000US-0583169.

XX 21-MAR-1996; 96US-014724P.

XX 21-MAR-1997; 97US-0821451.

XX 08-MAR-1999; 99US-0263810.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Yu G, Gentz R;

XX WPI: 2002-215019/27.

XX N-PSDB: ABL41783.

XX New antibody specific for human endometrial specific steroid-binding

XX PT factor (hESF) III, useful for detecting hESF III protein in biological

XX PT sample and to isolate or identify clones expressing the protein -

XX Disclosure; Fig 3; 36pp; English.

XX The present sequence represents a endometrial specific steroid-binding

XX CC factor (hESF) III. The full length protein has a molecular weight of

XX CC 8.10 Kda. The protein has homology to rat prostatic steroid-binding

XX CC protein C3. Antibodies which bind hESF proteins, such as hESF I, hESF II,

XX CC and hESF III are useful for isolating or to identify clones expressing

XX CC the polypeptides or to purify the polypeptides by affinity

XX CC chromatography. Agonists and antagonists of hESF proteins are useful

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLIMVLMALALLHCYADSGCKLLEDMEVKTIINSISPEYKELLOEFIDSDAAAEAMG 60
 DB 1 MKLIMVLMALALLHCYADSGCKLLEDMEVKTIINSISPEYKELLOEFIDSDAAAEAMG 60

OY 61 KFKOCFLNOSHRTLNKNGLMHHTVYDSIMCNMKS 95
 DB 61 KFKOCFLNOSHRTLNKNGLMHHTVYDSIMCNMKS 95

RESULT 9

ABBO9635
 ID ABBO9635 standard; Protein; 95 AA.

XX ABBO9635;

XX 29-MAY-2002 (first entry)

XX Human endometrial specific steroid-binding factor (hESF) III.

XX Human endometrial specific steroid-binding factor; ESF;

XX KW prostatic steroid-binding protein; hESF I; hESF II; hESF III; asthma.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT Peptide 1..21

XX FT Protein /note="signal peptide"

XX FT 22..95

XX FT /note="mature protein"

XX US6338948-B1.

XX 15-JAN-2002.

XX 30-MAY-2000; 2000US-0583169.

XX 21-MAR-1996; 96US-014724P.

XX 21-MAR-1997; 97US-0821451.

XX 08-MAR-1999; 99US-0263810.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Yu G, Gentz R;

XX WPI: 2002-215019/27.

XX N-PSDB: ABL41783.

XX New antibody specific for human endometrial specific steroid-binding

XX PT factor (hESF) III, useful for detecting hESF III protein in biological

XX PT sample and to isolate or identify clones expressing the protein -

XX Disclosure; Fig 3; 36pp; English.

XX The present sequence represents a endometrial specific steroid-binding

XX CC factor (hESF) III. The full length protein has a molecular weight of

XX CC 8.10 Kda. The protein has homology to rat prostatic steroid-binding

XX CC protein C3. Antibodies which bind hESF proteins, such as hESF I, hESF II,

XX CC and hESF III are useful for isolating or to identify clones expressing

XX CC the polypeptides or to purify the polypeptides by affinity

XX CC chromatography. Agonists and antagonists of hESF proteins are useful

OY 61 KFKOCFLNOSHRTLKNGFLMHTVYDSINCMKSN 95
 |||
 DB 61 KFKOCFLNOSHRTLKNGFLMHTVYDSINCMKSN 95

RESULT 10

AAV60038
 ID AAV60038 standard; Protein: 108 AA.

XX AAV60038;

XX 31-JAN-2000 (first entry)

DE Human endometrium tumour EST encoded protein 98.

KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
 KM treatment; uterine; gene therapy; expressed sequence tag.

XX Homo sapiens.

XX DEL9817948-A1.

XX 21-OCT-1999.

PF 17-APR-1998; 98DE-1017948.

XX 17-APR-1998; 98DE-1017948.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX MPI: 1999-591957/51.

DR N-PSDB: AAZ42013.

PT New nucleic acid sequences expressed in uterine cancer tissues, and
 PT derived polypeptides, for treatment of uterine and endometrial cancer
 and identification of therapeutic agents -

XX Claim 23: Page 314; 444pp; German.

CC This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (1) for recombinant
 CC expression of polypeptides (B) and (11) to isolate complete genes. (B)
 CC are used (1) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (11) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (111) for
 CC generation of specific antibodies. (A) are identified by assembling ESTs
 CC (expressed sequence tags) from a particular tissue type before comparison
 CC of expression patterns. This allows a significantly longer fragment of
 CC the gene to be revealed, so should reduce the number of failures
 CC associated with the fact that ESTs from different libraries may represent
 CC different parts of the same unknown gene, distorting the estimated
 CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in AAZ41981-Z42121.

XX SQ Sequence 108 AA:

Query Match 100.0%; Score 95; DB 20; Length 108;

Best Local Similarity 100.0%; Pred. No. 1.7e-89;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLLMVLMLALLHCYADSGCKLEDVMEKTSIDISPEYKELLOEFTDSAAAEAWG 60
 |||
 DB 14 MKLLMVLMLALLHCYADSGCKLEDVMEKTSIDISPEYKELLOEFTDSAAAEAWG 73

OY 61 KFKOCFLNOSHRTLKNGFLMHTVYDSINCMKSN 95
 |||
 DB 74 KFKOCFLNOSHRTLKNGFLMHTVYDSINCMKSN 108

RESULT 11
 AAM61649
 ID AAM61649 standard; peptide: 74 AA.

XX AAM61649;

XX 27-OCT-1998 (first entry)

DE Non-ocular disease marker 3.

KW Human; non-ocular disease; tear; cancer; breast; prostate.

XX Homo sapiens.

XX MO9835229-A1.

PD 13-AUG-1998.

PF 06-FEB-1998; 98WO-AU00071.

XX 07-FEB-1997; 97AU-0005009.

PA (MACQ-) MACQUARIE RES LTD.

PA (UNIX) UNISEARCH LTD.

PI Bolis S, Goolley AA, Herbert B, Molloy M, Morris C;

XX Walsh B, Wilcox M, Williams KL;

DR MPI: 1998-447373/38.

PT Screening for non-ocular disease - by analysing tears for marker
 PT proteins, particularly indicative of cancer and genetic disease,
 PT also new proteins and nucleic acid encoding them

XX Claim 6: Page 9; 14pp; English.

CC The markers AAM61647-W61649 are used for screening for, or detecting,
 CC non-ocular disease by analysing tears. Biochemicals, specifically
 CC proteins, are isolated from tears, particularly by chromatography or
 CC electrophoresis, especially two-dimensional polyacrylamide gel
 CC reagents, in (radio)immunoassay. The method is used to detect cancer,
 CC particularly of breast or prostate, or a genetic disease, in humans or
 CC animals.

XX SQ Sequence 74 AA:

Query Match 76.8%; Score 73; DB 19; Length 74;

Best Local Similarity 100.0%; Pred. No. 4.2e-67;

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 DSGCKLEDVMEKTSIDISPEYKELLOEFTDSAAAEAWGKFKOCFLNOSHRTLKNG 78
 |||
 DB 1 DSGCKLEDVMEKTSIDISPEYKELLOEFTDSAAAEAWGKFKOCFLNOSHRTLKNG 60

OY 79 LMMHTVYDSIWCN 91
 |||
 DB 61 LMMHTVYDSIWCN 73

RESULT 12

AAB51129
 ID AAB51129 standard; Peptide: 13 AA.

XX AAB51129;

XX 20-MAR-2001 (first entry)

DE Human mamaglobin peptide SEQ ID NO:29.

KW Human; mamaglobin; breast cancer; detection; diagnosis; antibody;
 KW vaccine; cytostatic; antimamaglobin.

```

XX OS Homo sapiens.
XX XX WO20007338-A1.
XX PN
XX XX
XX PD 07-DEC-2000.
XX XX
XX PF 26-MAY-2000; 2000WO-US14845.
XX XX
XX PR 28-MAY-1999; 99US-0136528.
XX PR 01-JUN-1999; 99US-0137048.
XX XX
XX PA (CORI-) CORIXA CORP.
XX PA (HEND/) HENDRICKSON R C.
XX PA (HONG/) HOUGHTON R L.
XX PA (REED/) REED S G.
XX PI
XX PI Fanger GR;
XX DR WPI; 2001-049928/06.
XX XX
XX PT Polypeptide comprising at least seven consecutive amino acid residues
XX PT of human mamaglobin, useful in the treatment and detection of breast
XX PT cancer -
XX PS
XX PS Example 4; Fig 10; 109pp; English.
XX XX
XX CC The present invention describes human mamaglobin peptides (1)
XX CC comprising at least 7 consecutive residues. Also described are: (1) a
XX CC vaccine comprising (1) with an immunostimulant which is an adjuvant;
XX CC (2) an isolated antibody (Ab1) or its antigen-binding fragment, which
XX CC specifically binds to a mamaglobin epitope having the sequence of
XX CC Pro2-3; (3) an isolated antibody (Ab2) or its antigen-binding fragment
XX CC that specifically binds to glycosylated mamaglobin; (4) a method for
XX CC inhibiting the development of breast cancer in a patient, comprising
XX CC administering (1) or Ab1 or Ab2; and (5) a method (M1) for determining
XX CC the presence or absence of breast cancer in a patient. (1) has
XX CC cytostatic activity. The polypeptides and antibodies are used in
XX CC vaccines for the prevention and treatment of breast cancer. They are
XX CC also used for diagnosis and monitoring of breast cancer. The present
XX CC sequence represents a human mamaglobin peptide sequence, which is
XX CC used in an example from the present invention.
XX SQ
XX SQ Sequence 13 AA:
XX
XX Query Match 12.6%; Score 12; DB 22; Length 13;
XX Best Local Similarity 100.0%; Pred. No. 2.9e-05;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MKLIMVLMAL 12
XX |||||||||
XX Db 1 MKLIMVLMAL 12
XX
XX RESULT 13
XX AAE07528
XX ID AAE07528 standard; Protein; 90 AA.
XX AC
XX AC AAE07528;
XX XX
XX DT 06-NOV-2001 (first entry)
XX XX
XX DE Human mamaglobin 207A 2 61551.2 protein.
XX XX
XX KM Human; lipophilin B; cytostatic; vaccine; mamaglobin 207A 2 61551.2;
XX KM gene therapy; uteroglobin; cancer; breast; ovary; prostate.
XX XX
XX OS Homo sapiens.
XX OS
XX PN WO200158947-A1.
XX XX
XX PD 16-AUG-2001.
XX XX

```

```

PF 08-FEB-2001; 2001WO-US04439.
XX XX
XX PR 11-FEB-2000; 2000US-0183495.
XX PR 28-JUN-2000; 2000US-0215735.
XX XX
XX PA (CORI-) CORIXA CORP.
XX PA
XX PI
XX PI Carter D, Vedvick TS, Vallieve-Douglass J, Houghton RL, Dillon DC;
XX XX
XX DR WPI; 2001-497069/54.
XX DR N-PSDB; AAD13771.
XX XX
XX PT Novel isolated complex two lipophilin-like polypeptides linked by at
XX PT least one disulfide bond, used to treat or prevent breast, ovarian or
XX PT prostate cancer -
XX PS
XX PS Example 5; Page 78-79; 91pp; English.
XX XX
XX CC The invention relates to a complex comprising a lipophilin-like
XX CC polypeptide linked by at least one disulfide bond to a second
XX CC lipophilin-like polypeptide. Lipophilin-like protein are members of
XX CC uteroglobin superfamily. Lipophilin-like proteins are useful in the
XX CC preparation of vaccines. The complex containing lipophilin-like
XX CC proteins are useful for treating or preventing breast, ovarian or
XX CC prostate cancer. The complex is also used for determining the
XX CC presence or absence of cancer in a patient, or monitor the progress
XX CC of cancer in a patient. Lipophilin DNA is also useful in gene therapy.
XX CC The present sequence is human mamaglobin 207A 2 61551.2 which is a
XX CC lipophilin-like protein.
XX SQ
XX SQ Sequence 90 AA:
XX
XX Query Match 12.6%; Score 12; DB 22; Length 90;
XX Best Local Similarity 100.0%; Pred. No. 0.00015;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MKLIMVLMAL 12
XX |||||||||
XX Db 1 MKLIMVLMAL 12
XX
XX RESULT 14
XX AAE07537
XX ID AAE07537 standard; Protein; 90 AA.
XX AC
XX AC AAE07537;
XX XX
XX DT 06-NOV-2001 (first entry)
XX XX
XX DE Human mamaglobin variant #8.
XX XX
XX KM Human; lipophilin B; cytostatic; vaccine; mamaglobin; variant; mutant;
XX KM gene therapy; uteroglobin; cancer; breast; ovary; prostate; ss.
XX XX
XX OS Homo sapiens.
XX OS
XX PN WO200158947-A1.
XX XX
XX PD 16-AUG-2001.
XX PD
XX PF 08-FEB-2001; 2001WO-US04439.
XX PF
XX PR 11-FEB-2000; 2000US-0183495.
XX PR 28-JUN-2000; 2000US-0215735.
XX XX
XX PA (CORI-) CORIXA CORP.
XX PA
XX PI
XX PI Carter D, Vedvick TS, Vallieve-Douglass J, Houghton RL, Dillon DC;
XX DR WPI; 2001-497069/54.
XX DR N-PSDB; AAD13787.
XX XX

```

PT Novel isolated complex two lipophilin-like polypeptides linked by at
PT least one disulfide bond, used to treat or prevent breast, ovarian or
PT prostate cancer. 1
XX
PS Example 6; Page 88; 91pp; English.
XX
CC The invention relates to a complex comprising a lipophilin-like
CC polypeptide linked by at least one disulfide bond to a second
CC lipophilin-like polypeptide. Lipophilin-like proteins are members of
CC uteroglobin superfamily. Lipophilin-like proteins are useful in the
CC preparation of vaccines. The complex containing lipophilin-like
CC proteins are useful for treating or preventing breast, ovarian or
CC prostate cancer. The complex is also used for determining the
CC presence or absence of cancer in a patient, or monitor the progress
CC of cancer in a patient. Lipophilin DNA is also useful in gene therapy.
CC The present sequence is human mammaglobin variant which is obtained
CC by deleting the amino acids 79-81. Mammaglobin is a lipophilin-like
CC protein.
XX
SQ Sequence 90 AA;

Query Match 12.6%; Score 12; DB 22; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLMLVLMAL 12
DB 1 MKLMLVLMAL 12

RESULT 15

AAM10179
ID AAM10179 standard; Protein: 93 AA.

XX AAM10179;

DT 12-AUG-1997 (first entry)

DE Mammary-specific secretory protein, mammaglobin.

KM mammaglobin; mammary-secretory protein; breast cancer; detection;
neoplastic disease; diagnosis.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..19

FT Protein /label= signal_peptide

FT /label= mature_protein

PN MO9638463-A1.

PD 05-DEC-1996.

PF 31-MAY-1996; 96MO-0508235.

PR 31-MAY-1995; 95US-0455896.

PA (UNIT) UNIV WASHINGTON.

PI Fleming TP, Watson MA;

DR WPT: 1997-034299/03.

DR N-PSDB; AAT50925.

XX Nucleic acid encoding mammary-specific secretory protein,
XX mammaglobin - used to develop prods. for the early diagnosis and
XX treatment of breast cancer neoplastic disease

PS Claim 3; Fig 2; 54pp; English.

XX The present sequence is that of a mammary-specific secretory protein

CC designated mammaglobin, which is overexpressed in 27% of stage I primary
CC breast cancer tumors. The anonymous sequence tag previously designated
CC DEST002 was used to demonstrate that mammaglobin is abundant in the
CC breast cancer tumor cell line MDA-MB-415. To isolate the full-length
CC mammaglobin cDNA (AAT50925), the mRNA was reverse transcribed from
CC this cell line and cloned using the RACE PCR technique. The nucleic acid
CC and protein can be used to develop prods. e.g. antibodies or probes, for
CC the detection and treatment of breast neoplastic disease.
XX

SQ Sequence 93 AA;

Query Match 12.6%; Score 12; DB 18; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLMLVLMAL 12
DB 1 MKLMLVLMAL 12

Search completed: January 25, 2003, 13:40:07
Job time : 36 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:38:25 ; Search time 31 Seconds
(without alignments)
631.435 Million cell updates/sec

Title: US-09-806-302A-2

Perfect score: 95
Sequence: 1 MKLMLVLMALHLHCYADS.....NFGLMHHTVYDSIMCKSN 95

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	8	8.4	249	4 Q9Y4U0	Q9Y4U0 homo sapien
2	8	8.4	256	16 Q9RUM5	Q9RUM5 deinococcus
3	8	8.4	372	8 Q956Q1	Q956Q1 acromys 1gn1
4	8	8.4	552	16 Q8X5S7	Q8X5S7 escherichia
5	8	8.4	937	12 Q9YP12	Q9YP12 citrus tatt
6	8	8.4	2105	12 Q9IR74	Q9IR74 apple stem
7	7	7.4	28	11 Q62677	Q62677 rattus norv
8	7	7.4	76	10 Q2A155	Q2A155 nicotiana t
9	7	7.4	92	3 Q9C2Q7	Q9C2Q7 neosporea
10	7	7.4	93	6 Q9GK61	Q9GK61 oryctolagus
11	7	7.4	107	16 Q9KN14	Q9KN14 vibrio chol
12	7	7.4	168	10 Q8M4T0	Q8M4T0 solanum phu
13	7	7.4	223	17 Q8TYO9	Q8TYO9 methanopyru
14	7	7.4	231	16 Q9AJX4	Q9AJX4 streptomyc
15	7	7.4	234	17 Q9HHY1	Q9HHY1 halobacteri
16	7	7.4	239	16 Q9RJV3	Q9RJV3 streptomyc

17	7	7.4	259	5 Q85MW6	Q85MW6 drosophila
18	7	7.4	269	2 Q937S2	Q937S2 bruceella me
19	7	7.4	291	10 Q9AWZ0	Q9AWZ0 oryza sativ
20	7	7.4	295	2 Q8VNT3	Q8VNT3 enterobacte
21	7	7.4	295	16 Q8YBR1	Q8YBR1 bruceella me
22	7	7.4	316	2 Q9KRW7	Q9KRW7 salmoneila
23	7	7.4	354	5 Q95V40	Q95V40 plasmodium
24	7	7.4	358	16 Q983D4	Q983D4 rhizobium 1
25	7	7.4	384	4 Q9GZNO	Q9GZNO homo sapien
26	7	7.4	384	11 Q9EPB7	Q9EPB7 mus musculu
27	7	7.4	384	11 Q9ESP4	Q9ESP4 rattus norv
28	7	7.4	402	5 Q18279	Q18279 caenorhabd1
29	7	7.4	425	5 Q26458	Q26458 drosophila
30	7	7.4	425	5 Q24338	Q24338 drosophila
31	7	7.4	469	5 Q461B3	Q461B3 drosophila
32	7	7.4	497	2 Q93PA1	Q93PA1 microscilla
33	7	7.4	525	11 Q9DC15	Q9DC15 mus musculu
34	7	7.4	568	10 Q94F19	Q94F19 arabidopsis
35	7	7.4	572	1 Q59663	Q59663 pyrodictium
36	7	7.4	576	16 Q9XA04	Q9XA04 streptomyc
37	7	7.4	590	6 Q02665	Q02665 oryctolagus
38	7	7.4	598	10 Q9C8T0	Q9C8T0 arabidopsis
39	7	7.4	605	16 Q92X67	Q92X67 rhizobium m
40	7	7.4	627	10 Q65380	Q65380 arabidopsis
41	7	7.4	665	16 Q97SC2	Q97SC2 streptococc
42	7	7.4	693	10 Q9FZB6	Q9FZB6 arabidopsis
43	7	7.4	758	16 Q910L8	Q910L8 pseudomonas
44	7	7.4	767	10 Q94AJ9	Q94AJ9 arabidopsis
45	7	7.4	769	5 Q95TJ0	Q95TJ0 drosophila

ALIGNMENTS

RESULT 1	ID	Q9Y4U0	PRELIMINARY:	PRT:	249 AA.
AC	Q9Y4U0				
DT	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DE	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Hypothetical 29.1 kDa protein (Fragment).				
GN	DKFZP564I052.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]	SEQUENCE FROM N.A.			
RC	TISSUE=BRIN;				
RA	Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;				
RU	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.				
DR	EMBL: AL080063; CAB45694.1; -				
DR	InterPro: IPR001005; Myb_DNA-binding.				
DR	InterPro: IPR000433; Znf_ZZ.				
DR	Pfam: PF00249; myb_DNA-binding; 1.				
DR	Pfam: PF00569; Zz; 1.				
DR	SMART: SM00395; SANT; 1.				
DR	SMART: SM00291; Znf_ZZ; 1.				
DR	PROSITE: PS50090; MYB_3; 1.				
DR	PROSITE: PS01357; ZF_ZZ_1; UNKNOWN_1.				
KW	Hypothetical protein; Nuclear protein.				
FT	NON_TER				
SQ	SEQUENCE	249 AA; 29104 MW; DB887EA33FC32 CRC64;			
Query Match		8.4%; Score 8; DB 4; Length 249;			
Best Local Similarity		100.0%; Pred No. 4.6;			
Matches	8; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	40 PEYKELQ 47				

```

Db      136 PEYKELLQ 143

RESULT 2
Q9RUM5 PRELIMINARY: PRT; 256 AA.
AC Q9RUM5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Outer membrane protein.
GN DRI359.
OS Delnoccocus radiodurans.
OC Bacteria; Thermus/Delnoccocus group; Delnoccoci; Delnoccocales;
OC Delnoccocaceae; Delnoccoccus.
ON NCBI_TaxID=1259;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RA MEDLINE=20036696; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.D., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Campbell W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischman R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Delnoccoccus
RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001982; AAF10931.1; -.
DR TIGR; DR1359; -.
DR InterPro; IPR004872; Lipoprotein_9.
DR Pfam; PF03180; Lipoprotein_9; 1.
KW Complete proteome.
SQ
SEQUENCE 256 AA: 27737 MW: CC72A304E1B3D0D CRC64;

Query Match      8.4%; Score 8; DB 16; Length 256;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 LMLALLL 14
       11111111
Db      5 LMLALLL 12

RESULT 3
Q956Q1 PRELIMINARY: PRT; 372 AA.
AC Q956Q1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cytochrome b (fragment).
GN CYTB OR CYTOCHROME B.
OS Acomyx ignitus (flying squirrel mouse).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Acomys.
ON NCBI_TaxID=60745;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BUSHACKERS;
RA MEDLINE=98334580; PubMed=9668005;
RA Barome P.O., Monnerot M., Gautun J.C.;
RT "Intra generic phylogeny of Acomys (Rodentia, Muridae) using
RT mitochondrial gene cytochrome b.";
RL Mol. Phylogene. Evol. 9:560-566(1998).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

```

```

CC      BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR EMBL; Z96064; CAB09438.1; -.
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c1; 1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 372
SQ SEQUENCE 372 AA: 41903 MW: 387EE9BAD5B94C6E CRC64;

Query Match      8.4%; Score 8; DB 8;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 LMLALLL 14
       11111111
Db      234 LMLALLL 241

RESULT 4
Q8X5S7 PRELIMINARY: PRT; 552 AA.
AC Q8X5S7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Formate-dependent nitrite reductase, possible assembly function
DE (Cytochrome c-type biogenesis protein NrfE).
GN NrfE OR Z5673 OR EC55056.
OS Escherichia coli O157:H7.
OC Escherichia.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=83334;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grothbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005640; AAG59272.1; -.
DR EMBL; AP002568; BAB38479.1; -.
DR InterPro; IPR002541; CytC_asm.
DR InterPro; IPR003568; CytC_biol_CcmF.
DR InterPro; IPR003567; CytC_biol.
DR Pfam; PF01578; CytC_asm; 1.
DR PRINTS; PR01410; CCBIOGENESIS.
DR TIGRfams; TIGR00353; nrfE; 1.
KW Complete proteome.
SQ SEQUENCE 552 AA: 61368 MW: 849838B944EDB8AD CRC64;

Query Match      8.4%; Score 8; DB 16; Length 552;

```


Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LMLALLL 14
|||||
Db 114 LMLALLL 121

RESULT 5

Q9YPI2 PRELIMINARY; PRT; 937 AA.
AC Q9YPI2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF1 (Fragment).
OS Citrus tatter leaf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Capilliovirus.
OX NCBI_TaxID=33759;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103780; PubMed=8277280;
RA Yoshikawa N., Imaizumi M., Takahashi T., Inoue N.;
RT "Striking similarities between the nucleotide sequence and genome organization of citrus tatter leaf and apple stem grooving capillioviruses.";
RT J. Gen. Virol. 74:2743-2747(1993).
RL EMBL; D16368; BAA03869.1; -
DR Interpro: IPR001788; RNA_dep_RNApol2. 1.
DR Pfam: PF00978; RNA_dep_RNApol2. 1.
FT NON_TER
SQ SEQUENCE 937 AA; 108705 MW; 286DB49F829FAA3B CRC64;

Query Match 8.4%; Score 8; DB 12; Length 937;
Best Local Similarity 100.0%; Pred. No. 14;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KELLQEFI 50
|||||
Db 55 KELLQEFI 62

RESULT 6
Q9IR74 PRELIMINARY; PRT; 2105 AA.
AC Q9IR74;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 241K polypeptide.
OS Apple stem grooving virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Capilliovirus.
OX NCBI_TaxID=28347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LI-23;
RA Terauchi H., Magome H., Yoshikawa N., Takahashi T.;
RT "Nucleotide sequence of the genome of apple stem grooving capilliovirus RT 1 isolate and construction of an infectious cDNA clone of the genome containing a cauliflower mosaic virus 35S RNA.";
RL Ann. Phytopathol. Soc. Jpn. 63:432-436(1997).
DR EMBL; AB004063; BAA98054.1; -
DR MEROPS; C35.001; -
DR Interpro: IPR002173; PfkB.
DR Interpro: IPR001788; RNA_dep_RNApol2.
DR Interpro: IPR000606; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNApol2. 1.
DR Pfam: PF01443; Viral_helicase1. 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
KM Polypeptidein.
SQ SEQUENCE 2105 AA; 241787 MW; ABA98A31633A1A4F CRC64;

Query Match 8.4%; Score 8; DB 12; Length 2105;

Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KELLQEFI 50
|||||
Db 1223 KELLQEFI 1230

RESULT 7

Q62677 PRELIMINARY; PRT; 28 AA.
AC Q62677; 009145; 009065;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Alpha-1 collagen type IV (Fragment).
GN COL4A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97107443; PubMed=8950183;
RA Grande J.P., Weider D.C., Kluge D.L., Wieben E.D.;
RT "Structure of the rat collagen IV promoter.";
RT Biochim. Biophys. Acta 1309:85-88(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Grande J.P., Haugen J.D.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85606; AAB47426.1; -
KM Collagen.
FT NON_TER
SQ SEQUENCE 28 AA; 3113 MW; 96DFC6C741E07809 CRC64;

Query Match 7.4%; Score 7; DB 11; Length 28;
Best Local Similarity 100.0%; Pred. No. 7.7;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LAALLH 15
|||||
Db 13 LAALLH 19

RESULT 8
O24155 PRELIMINARY; PRT; 76 AA.
AC O24155;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TA20.
GN TA20.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ANTHER;
RA Beals T.P., Goldberg R.B.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 53-76 FROM N.A.
RC TISSUE=ANTHER;
RA Koltunow A.M., Treutner J., Cox K.H., Wallroth M., Goldberg R.B.;
RL Plant Cell 2:1201-1224(1990).
RN [3]
RP SEQUENCE OF 53-76 FROM N.A.
RC TISSUE=ANTHER;
RA Beals T.P., Goldberg R.B.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U73164; AAB18190.1; -
 DR EMBL: U73165; AAB18191.1; -
 SQ SEQUENCE 76 AA; 8269 MW; 2F0A47BF7A52F7CC CRC64;

Query Match 7.4%; Score 7; DB 10; Length 76;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LMVLM 10
 DB 19 LMVLM 25

RESULT 9

O9C207 PRELIMINARY; PRT; 92 AA.

AC O9C207;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 10.6 kDa protein.
 GN 104H10.170.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Allyn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL513410; CAC28583.2; -
 KW Hypothetical protein.
 SQ SEQUENCE 92 AA; 10642 MW; 548FBB43ABD72E98 CRC64;

Query Match 7.4%; Score 7; DB 3; Length 92;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KLMLM 8
 DB 58 KLMLM 64

RESULT 10

O9GK61 PRELIMINARY; PRT; 93 AA.

AC O9GK61;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Lipophilin CS.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-SUBMAXILLARY;
 RC Zhao C., Nguyen T.X., Lehrer R.I.;
 RT "Rabbit Lipophilin";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF308620; AAG42808.1; -
 DR InterPro: IPR003627; Mamgb/prostatn.
 DR InterPro: IPR000329; Uteroglobln.
 DR Pfam: PF01099; Uteroglobln; 1.
 DR PRINTS: PR00486; UTERGLOBLN.
 DR ProDom: PD029354; Mamgb/prostatn; 1.

DR SMART: SM00096; UMG; 1.
 SQ SEQUENCE 93 AA; 10609 MW; E2B015AD319B2249 CRC64;

Query Match 7.4%; Score 7; DB 6; Length 93;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VMLMAL 12
 DB 6 VMLMAL 12

RESULT 11

O9KN14 PRELIMINARY; PRT; 107 AA.

AC O9KN14;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein VCA0152.
 GN VCA0152.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,
 RA McDonald L., Uitterlind T., Fleischmann R.D., Niemann W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RL Nature 406:477-483(2000).
 DR EMBL: AE004356; AAF96065.1; -
 DR TIGR: VCA0152; -
 DR InterPro: IPR005133; Phag_MnhG_YufB.
 DR Pfam: PF03334; Phag_MnhG_YufB; 1.
 DR TIGRFAMS: TIGR01300; CPA3_mnhg_phag; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 107 AA; 11488 MW; 3B0F0CC77BF2D9B8 CRC64;

Query Match 7.4%; Score 7; DB 16; Length 107;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 MLMAALL 14
 DB 3 MLMAALL 9

RESULT 12

O8W4T0 PRELIMINARY; PRT; 168 AA.

AC O8W4T0;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Resistance gene analog (Fragment).
 OS Solanum phureja x Solanum stenotomum.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=176295;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BD410-33;
 RA Simko I., Jones R.W.;
 RT "Mapping of Resistance Gene Analogs in Potato.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY059428; AL30114.1; -
 DR InterPro: IPR002182; NB-ARC.
 DR Pfam: PF00931; NB-ARC; 1.
 FT NON_TER 1 168
 FT NON_TER 1 168
 SQ SEQUENCE 168 AA; 19000 MW; DB34AF796AE74EC4 CRC64;

Query Match 7.4%; Score 7; DB 10; Length 168;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 KELLOEF 49
 Db 41 KELLOEF 47

RESULT 13

ID 08TY09 PRELIMINARY; PRT; 223 AA.
 AC 08TY09;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Predicted dinucleotide-binding enzyme.
 GN MK0233.
 OS Methanopyrus kandleri.
 OC Archaea: Euryarchaeota: Methanopyri: Methanopyrales: Methanopyraceae;
 OC Methanopyrus.
 NCBI_TaxID=2320;
 RX MEDLINE=21927647; PubMed=11930014;
 RP SEQUENCE FROM N.A.
 RC STRAIN-AV19 / DSM 6324 / JCM 9639;
 RA Slesarev A.T., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Stuchebkina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Nalete D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Mal'lykh A.G., Koonin E.V., Kozhavkin S.A.;
 RT "The complete genome of hyperthermophilic Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL: AE010321; AM01450.1; -
 KW Complete proteome.
 SQ SEQUENCE 223 AA; 23975 MW; 3C379224E4AF6F92 CRC64;

Query Match 7.4%; Score 7; DB 17; Length 223;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 SDAAAEA 58
 Db 64 SDAAAEA 70

RESULT 14

ID 09AJX4 PRELIMINARY; PRT; 231 AA.
 AC 09AJX4;
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Putative oxidoreductase.
 GN SCO1803 OR SC133.02.
 OS Streptomyces coelicolor.
 OC Bacteria: Firmicutes: Actinobacteria: Actinobacteriales;
 OC Actinomycetales: Streptomycetales: Streptomycetaceae: Streptomyces.
 NCBI_TaxID=1902;
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Harris D.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RP [3]
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=843436;
 RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.

RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 DR EMBL: AL513407; CAC28543.1; -
 DR HSSP: 064105; 10AA.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRfamily.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN.1.
 KW Oxidoreductase.
 SQ SEQUENCE 231 AA; 23789 MW; 6AEDBA0306B94086 CRC64;

Query Match 7.4%; Score 7; DB 16; Length 231;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 SDAAAEA 58
 Db 139 SDAAAEA 145

RESULT 15

ID 09HHY1 PRELIMINARY; PRT; 234 AA.
 AC 09HHY1;
 DT 01-MAR-2001 (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE TRK potassium uptake system protein.
 GN TRKA2 OR VNG6175G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea: Euryarchaeota: Halobacteria: Halobacteriales;
 OC Halobacteriaceae: Halobacterium.
 NCBI_TaxID=64091;
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shroya J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Wells R., Goo Y.A.,
 RA Leitner B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasarma S.;

RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005152; AG20842.1; -.
DR InterPro: IPR000309; TrkA_Kuptake.
DR InterPro: IPR003148; TrkA_N.
DR Pfam: PF02080; TrkA-C; 1.
DR Pfam: PF02254; TrkA-N; 1.
DR PRINTS: PR00335; KUPTAKETRKA.
KW Plasmid; Complete proteome.
SQ SEQUENCE 234 AA; 25291 MW; 92D50B3E557ADFD CRC64;

Query Match 7.4%; Score 7; DB 17; Length 234;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LMVLMIA 10
DB 80 LMVLMIA 86

Search completed: January 25, 2003, 13:41:17
Job time : 34 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 10:27:54 ; Search time 306 seconds
(without alignments)
3503.108 Million cell updates/sec

Title: US-09-806-302A-1

Perfect score: 476

Sequence: 1 acgagctgcacgcagcagcact.....cctcaattcattcattca 476

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	476	100.0	476	18	AAT94832 Human endometrial
2	476	100.0	476	21	AAA5730 Human endometrial
3	476	100.0	476	21	AAA09104 Human endometrial
4	476	100.0	476	22	AAF25214 cDNA encoding of a
5	476	100.0	476	24	ABL41783 Human endometrial
6	433.2	91.0	496	24	AAK9492 cDNA of human lipo
7	431.6	90.7	517	20	AAZ42013 Human endometrium
8	431.6	90.7	517	21	AAA09118 Clone Mamm-X cDNA,
9	424	89.1	473	21	AAZ45008 Human 5' EST, Isola

10	408.4	85.8	485	20	AAK36138 DNA encoding a hum
11	407.4	85.6	491	24	ABLe7870 Ovary cancer relat
12	295	62.0	333	22	AAI20172 Human breast cancer
13	291.8	61.3	333	22	AAI17037 Human breast cancer
14	291.8	61.3	370	22	AAI08758 Human breast cancer
15	280	58.8	368	24	ABU8350 Human breast cancer
16	223.8	47.0	495	19	AAVA1580 Human ovarian can
17	223.8	47.0	495	24	ABK89127 Nucleotide sequenc
18	223.8	47.0	495	24	AAAD2632 cDNA #2 encoding h
19	223.8	47.0	503	18	AAAT50925 Human steroid bind
20	223.8	47.0	503	19	AAV17905 cDNA encoding mamm
21	223.8	47.0	503	20	AAAT6966 Mammaglobin gene S
22	223.8	47.0	503	21	AAAI2632 cDNA encoding the
23	223.8	47.0	503	22	AAAD13755 Human mammaglobin
24	223.8	47.0	535	21	AAAG4845 Human mammaglobin
25	223.8	47.0	548	19	AAV17906 Mammaglobin gene S
26	222.2	46.7	503	22	AAI17266 Mammaglobin cDNA S
27	219.6	46.1	820	22	AAI19890 Human breast cancer
28	217	45.6	801	22	AAI10310 Human breast cancer
29	211.6	44.5	429	22	AAAD13786 Human mammaglobin
30	210	44.1	429	22	AAAD13783 Human mammaglobin
31	208.4	43.8	429	22	AAAD13780 Human mammaglobin
32	208.4	43.8	429	22	AAAD13781 Human mammaglobin
33	208.4	43.8	429	22	AAAD13782 Human mammaglobin
34	208.4	43.8	429	22	AAAD13785 Human mammaglobin
35	205.6	43.2	388	22	AAAD13773 Human mammaglobin
36	204.4	42.9	408	22	AAI11273 Human breast cancer
37	202.4	42.5	388	22	AAAD13772 Human mammaglobin
38	202.4	42.5	388	22	AAAD13774 Human mammaglobin
39	202.4	42.5	388	22	AAAD13775 Human mammaglobin
40	202.4	42.5	388	22	AAAD13776 Human mammaglobin
41	202.4	42.5	388	22	AAAD13777 Human mammaglobin
42	202.4	42.5	388	22	AAAD13778 Human mammaglobin
43	202.4	42.5	388	22	AAAD13779 Human mammaglobin
44	201.2	42.3	403	21	AAAI2635 PCR amplified 403
45	195.2	41.0	467	22	AAH5518 Human breast tumor

ALIGNMENTS

RESULT 1
AAT94832 standard; DNA: 476 BP.

ID	AAT94832	standard; DNA: 476 BP.
XX	AC	AAT94832:
XX	XX	
XX	XX	27-MAR-1998 (first entry)
DE	XX	Human endometrial specific steroid-binding factor III DNA.
XX	XX	
KW	KW	Endometrial specific steroid-binding factor III; ESF III; human;
KW	KW	endometrium; phospholipase A2 inhibitor; polychlorinated biphenyl;
KW	KW	antiaggregant; inflammation; asthma; rhinitis; cystic fibrosis;
KW	KW	airway disease; neoplasia; atopy; therapy; diagnosis; ss.
XX	XX	
OS	XX	Homo sapiens.
XX	XX	
PH	FT	Key
FT	FT	CDS
FT	FT	sig_peptide
FT	FT	mat_peptide
FT	FT	
XX	XX	
XX	XX	W09734997-A1.
PN	XX	25-SEP-1997.
XX	XX	
PD	XX	21-MAR-1996; 96WO-US03857.
PF	XX	
XX	XX	21-MAR-1996; 96WO-US03857.
PR	XX	

Location/Qualifiers

/*tag= a
46..108
/*tag= b
109..330
/*tag= c


```

Db 1 ACGAGCTGCGACGACGACTGGAACAGACAGCCGCTCGCATGAACTGCTGATG 60
Oy 61 GTCTCATGCTGGCGCCCTCCTCCGACACTGTATGAGATTTCGGCGCAAACTCCG 120
Db 61 GTCTCATGCTGGCGCCCTCCTCCGACACTGTATGAGATTTCGGCGCAAACTCCG 120
Oy 121 GAGGACATGTTGTAAGAAAGACATCAATTCGACATATCATATCAATCAAGAGCTT 180
Db 121 GAGGACATGTTGTAAGAAAGACATCAATTCGACATATCATATCAATCAAGAGCTT 180
Oy 181 CTTCAGAGATTATAGACAGTATGCGCGTGCAGAGGCTATGGGAATTCAGCAGTGT 240
Db 181 CTTCAGAGATTATAGACAGTATGCGCGTGCAGAGGCTATGGGAATTCAGCAGTGT 240
Oy 241 TTCTCAACACGATGACATGAGACTGTGAAGAACTTGGCTGATGATGATGAGTAC 300
Db 241 TTCTCAACACGATGACATGAGACTGTGAAGAACTTGGCTGATGATGATGAGTAC 300
Oy 301 GACAGCATTGGTGTATATATGAAGATTAATTAACCTTACCAAGCGGTTGGCTCAGAG 360
Db 301 GACAGCATTGGTGTATATATGAAGATTAATTAACCTTACCAAGCGGTTGGCTCAGAG 360
Oy 361 GCTACGACTATGCGCAGACACTCATCTGTTGATGCTAGAAACCACTTCTTCTGTGT 420
Db 361 GCTACGACTATGCGCAGACACTCATCTGTTGATGCTAGAAACCACTTCTTCTGTGT 420
Oy 421 GCTTTTATGTGGGAACCTGTAGAACACTGTGAAACCTCAATTCATTCATTTC 476
Db 421 GCTTTTATGTGGGAACCTGTAGAACACTGTGAAACCTCAATTCATTCATTTC 476

RESULT 3
AAA09104
ID AAA09104 standard: DNA; 476 BP.
XX
AC AAA09104:
XX
DT 10-AUG-2000 (first entry)
XX
DE Human endometrial specific steroid binding factor coding sequence.
XX
KW Endometrial specific steroid binding factor; ESBPII; diagnosis;
KW gynecological cancer; uterine; breast; endometrial; ovarian;
KW antibody; gene therapy; cytostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..333
FT /tag= a
XX
PN WO200020044-A1.
XX
PD 13-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22753.
XX
PR 02-OCT-1998; 98US-0102743.
XX
PA (DIND-) DIADEXUS- LLC.
XX
PI Macina RA:
XX
OS
XX
DR MPI: 2000-303649/26.
XX
DR P-PSDB; AAY92226.
XX
XX
PT Diagnosing, staging and monitoring gynecological cancer comprising
PT using an elevated level of ESBPII in a patient as an indicator of
PT cancer
XX
PS Claim 6; Page 28; 32pp; English.
XX

```

```

CC The levels of human endometrial specific steroid binding factor
CC (ESBPII) can be measured and compared to control levels and used to
CC diagnose the presence of a gynecological (uterine, breast, endometrial,
CC or ovarian) cancer in a patient. ESBPII levels can also be used to
CC diagnose metastasis, to stage or monitor gynecological cancer.
CC Antibodies specific for ESBPII can be used to treat gynecological
CC cancers.
XX
SQ Sequence 476 BP; 131 A; 115 C; 105 G; 125 T; 0 other:

Query Match 100.0%; Score 476; DB 21; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.0e-136;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ACGAGCTGCGACGACGACTGGAACAGACAGCCGCTCGCATGAACTGCTGATG 60
Db 1 ACGAGCTGCGACGACGACTGGAACAGACAGCCGCTCGCATGAACTGCTGATG 60
Oy 61 GTCTCATGCTGGCGCCCTCCTCCGACACTGTATGAGATTTCGGCGCAAACTCCG 120
Db 61 GTCTCATGCTGGCGCCCTCCTCCGACACTGTATGAGATTTCGGCGCAAACTCCG 120
Oy 121 GAGGACATGTTGTAAGAAAGACATCAATTCGACATATCATATCAAGAGCTT 180
Db 121 GAGGACATGTTGTAAGAAAGACATCAATTCGACATATCATATCAAGAGCTT 180
Oy 181 CTTCAGAGATTATAGACAGTATGCGCGTGCAGAGGCTATGGGAATTCAGCAGTGT 240
Db 181 CTTCAGAGATTATAGACAGTATGCGCGTGCAGAGGCTATGGGAATTCAGCAGTGT 240
Oy 241 TTCTCAACACGATGACATGAGACTGTGAAGAACTTGGCTGATGATGATGAGTAC 300
Db 241 TTCTCAACACGATGACATGAGACTGTGAAGAACTTGGCTGATGATGAGTAC 300
Oy 301 GACAGCATTGGTGTATATATGAAGATTAATTAACCTTACCAAGCGGTTGGCTCAGAG 360
Db 301 GACAGCATTGGTGTATATATGAAGATTAATTAACCTTACCAAGCGGTTGGCTCAGAG 360
Oy 361 GCTACGACTATGCGCAGACACTCATCTGTTGATGCTAGAAACCACTTCTTCTGTGT 420
Db 361 GCTACGACTATGCGCAGACACTCATCTGTTGATGCTAGAAACCACTTCTTCTGTGT 420
Oy 421 GCTTTTATGTGGGAACCTGTAGAACACTGTGAAACCTCAATTCATTTCATTTC 476
Db 421 GCTTTTATGTGGGAACCTGTAGAACACTGTGAAACCTCAATTCATTTCATTTC 476

RESULT 4
AAF25214
ID AAF25214 standard: cDNA; 476 BP.
XX
AC AAF25214:
XX
DT 30-APR-2001 (first entry)
XX
DE cDNA encoding of an endometrial specific steroid binding factor III.
XX
KW Human: endometrial specific steroid binding factor; hESF; hESFI;
KW hESFII; inflammation; asthma; rhinitis; cystic fibrosis; airway disease;
KW neoplasia; atopy; phospholipase A2; polychlorinated biphenyl; chemotaxis;
KW phagocytosis; platelet aggregation; eicosanoid; endometrial cell; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..333
FT /tag= a
FT /product= "endometrial specific steroid binding
FT sig_peptide factor (hESF) III"
FT /tag= b
XX
PS US6174992-B1.
XX

```


QY 66 CATGCTGGGCGCCCTCCCTCCCTGCTATGCAAGATTCTGGCTGCAAACTCCTGGAGGA 125
DB 85 CATGCTGGGCGCCCTCCCTCCCTGCTATGCAAGATTCTGGCTGCAAACTCCTGGAGGA 144
QY 126 CATGGTTGAAAAGACCATTCAATTCGACATATCTATACCTGAATACAAAGACCTTCTTCA 185
DB 145 CATGGTTGAAAAGACCATTCAATTCGACATATCTATACCTGAATACAAAGACCTTCTTCA 204
QY 186 AAGATTCAATAGACATGATGCGCGTGCAGAGGCTATGGGAAATTCACACAGTGTTCCT 245
DB 205 AAGATTCAATAGACATGATGCGCGTGCAGAGGCTATGGGAAATTCACACAGTGTTCCT 264
QY 246 CAACCAAGTCACATAGAACTCTGAAAAAATCTTGGAGTATGATGATACAGTATAGACAG 305
DB 265 CAACCAAGTCACATAGAACTCTGAAAAAATCTTGGAGTATGATGATACAGTATAGACAG 324
QY 306 CATTTGGTGTATATGAGAGATTAATTACTTTACCCAGGCGTTGGCTCAGAGGCTTAC 365
DB 325 CATTTGGTGTATATGAGAGATTAATTACTTTACCCAGGCGTTGGCTCAGAGGCTTAC 384
QY 366 AAGACTATGGCCAGAACTCATCTGTGATGCTAGAAACCACTTCTCTTGTG---TTGC 422
DB 385 AAGACTATGGCCAGAACTCATCTGTGATGCTAGAAACCACTTCTCTTGTG---TTGC 444
QY 423 TTTTATGTGGGAAGTGTGAGCAAACTGTTGAACCTCAATT 464
DB 445 TTTTATGTGGGAAGTGTGAGCAAACTGTTGAACCTCAATT 486

RESULT 9
AAZ43008
ID AAZ43008 standard; cDNA; 473 BP.
XX

AC AAZ43008;
XX
DT 01-FEB-2000 (first entry)
XX

DE Human 5' EST isolated from a cDNA library SEQ ID NO:767.
XX

KM Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KM gene therapy; chromosome mapping; upstream regulatory sequence;
KM forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX

OS Homo sapiens.
XX

XX
PN W09953051-A2.
XX

PD 21-OCT-1999.
XX

PF 09-APR-1999; 99WO-1B00712.
XX

PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX

XX (GEST) GENSET.
PA

PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX

DR WPI; 2000-038446/03.
XX

DR P-PSDB; AAY65394.
XX

PT Novel secreted protein 5' expressed sequence tag sequences used in
XX diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX

PS Claim 1: Page 569; 837pp; English.
XX

CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43075. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated

CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX

Sequence 473 BP; 130 A; 113 C; 104 G; 123 T; 3 other:

Query Match 89.1%; Score 424; DB 21; Length 473;
Best Local Similarity 96.9%; Pred. No. 1; Be-120;

Matches 440; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

QY 6 CTGCCACGACGACGTGACACAGACAGCCGCTCCGATGAAGCTGATGCTCT 65
DB 20 CTGCCACGACGACGTGACACAGACAGCCGCTCCGATGAAGCTGATGCTCT 79
QY 66 CATGCTGGGCGCCCTCCCTCCCTGCTATGCAAGATTCTGGCTGCAAACTCCTGGAGGA 125
DB 80 CATGCTGGGCGCCCTCCCTCCCTGCTATGCAAGATTCTGGCTGCAAACTCCTGGAGGA 139

QY 126 CATGGTTGAAAAGACCATTCAATTCGACATATCTATACCTGAATACAAAGACCTTCTTCA 185
DB 140 CATGGTTGAAAAGACCATTCAATTCGACATATCTATACCTGAATACAAAGACCTTCTTCA 199
QY 186 AAGATTCAATAGACATGATGCGCGTGCAGAGGCTATGGGAAATTCACACAGTGTTCCT 245
DB 200 AAGATTCAATAGACATGATGCGCGTGCAGAGGCTATGGGAAATTCACACAGTGTTCCT 259

QY 246 CAACCAAGTCACATAGAACTCTGAAAAAATCTTGGAGTATGATGATACAGTATAGACAG 305
DB 260 CAACCAAGTCACATAGAACTCTGAAAAAATCTTGGAGTATGATGATACAGTATAGACAG 319
QY 306 CATTTGGTGTATATGAGAGATTAATTACTTTACCCAGGCGTTGGCTCAGAGGCTTAC 365
DB 320 CATTTGGTGTATATGAGAGATTAATTACTTTACCCAGGCGTTGGCTCAGAGGCTTAC 379

QY 366 AAGACTATGGCCAGAACTCATCTGTGATGCTAGAAACCACTTCTCTTGTG---TTGC 422
DB 380 AAGACTATGGCCAGAACTCATCTGTGATGCTAGAAACCACTTCTCTTGTG---TTGC 439
QY 423 TTTTATGTGGGAAGTGTGAGCAAACTGTTGAACCTCAATT 466
DB 440 TTTTATGTGGGAAGTGTGAGCAAACTGTTGAACCTCAATT 473

RESULT 10
AAZ36138
ID AAZ36138 standard; DNA; 485 BP.
XX

AC AAZ36138;
XX

DT 26-JUL-1999 (first entry)
XX

DE DNA encoding a human mammothobin homologue (HMH).
XX

KW Human mammothobin homologue; HMH; antagonist; neoplastic disorder;
KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
KW teratocarcinoma; endometriosis; ss.
XX

OS Homo sapiens.
XX

XX
FH Key Location/Qualifiers
FT CDS 43..330

FT /*tag= a
FT /product= "mammoglobin homologue"
PN WO919487-A1.
XX
XX 22-APR-1999.
PD
XX 14-OCT-1998; 98WO-US21729.
XX
XX 16-OCT-1997; 97US-0951750.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Hillman JL, Murry LE, Shah P;
XX
XX WPI: 1999-302531/25.
DR P-PSDB; AAY025590.
XX
XX New human mammoglobin homolog (HMH), useful for diagnosing, treating
PT or preventing disorders associated with expression of HMH
XX
XX
PS Claim 7; Fig 1A-B; 63pp: English.
XX
XX The present sequence encodes a human mammoglobin homologue (HMH).
CC Antagonists of the HMH polypeptide can be used to treat neoplastic
CC disorders including adenocarcinoma, leukemia, lymphoma, melanoma,
CC myeloma, sarcoma and teratocarcinoma. A vector expressing the
CC complement of the polynucleotide encoding HMH may be administered
CC to a subject to treat or prevent neoplastic disorders or endometriosis.
CC Antibodies which bind HMH may also be used in the diagnosis of
CC conditions or diseases characterized by expression of HMH, or in assays
CC to monitor patients being treated with HMH, agonists, antagonists or
CC inhibitors. Polynucleotides encoding HMH may also be used diagnostically
CC to detect and quantitate gene expression in biopsied tissues. With
CC respect to cancer a relatively high amount of transcript may indicate a
CC predisposition for the development of disease. The nucleic acid sequences
CC which encode HMH may also be used to generate hybridization probes useful
CC for mapping the naturally occurring genomic sequence. HMH, and its
CC fragments/variants can be used for screening libraries of compounds in
CC drug screening techniques.
XX
XX
SQ Sequence 485 BP; 136 A; 116 C; 102 G; 129 T; 2 other:

Query Match 85.8%; Score 408.4; DB 20; Length 485;
Best Local Similarity 95.2%; Pred. No. 1.2e-115;
Matches 455; Conservative 0; Mismatches 16; Indels 7; Gaps 3;

QY 6 CTGCCACGACGACGTGAACACAGACAGCGCCCTCGCATGAAGCTCTGATGTCCT 65
DB 3 CTGCCACGACGACGTGAACACAGACAGCGCCCTCGCATGAAGCTCTGATGTCCT 62
QY 66 CATGCTGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125
DB 63 CATGCTGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
QY 126 CATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 185
DB 123 CATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 182
QY 186 AGAGTTTATAGACATGATGCGGCTGACAGAGCTATGGGAAATTCAGCACTGTTTCT 245
DB 183 AGAGTTTATAGACATGATGCGGCTGACAGAGCTATGGGAAATTCAGCACTGTTTCT 242
QY 246 CAACGAGTCATAGAACTCTGAAAAAAGTTTGACATGATGATGATGATGATGATGATG 305
DB 243 CAACGAGTCATAGAACTCTGAAAAAAGTTTGACATGATGATGATGATGATGATGATG 302
QY 306 CATTTGGTGTATATGAGAGTAATTAATTACCCAGGCGTTTGGCTCAGAGGCGTAC 365
DB 303 CATTTGGTGTATATGAGAGTAATTAATTACCCAGGCGTTTGGCTCAGAGGCGTAC 362
QY 366 AGAGTATGCG---CCAGAACTCATCTGTGATTGCTAGAAACCATTTCTTCTTGTG---T 419
XX
XX

DB 363 AGACTAATGCCCCAGAACTCATCTGTTGATTGCTAGAAACCATTTCTTCTTGTGTT 422
QY 420 TCGTTTTTATGTGGGAGCTAGACAACTGTTGAACCTC-AATTCAATTCATTTC 476
DB 423 GTCTTTTTATGTGGGAGCTAGACAACTGTTGAACCTC-AATTCAATTCATTTC 480

RESULT 11
ABL67870
ID ABL67870 standard; DNA; 491 BP.
XX
XX ABL67870;
XX
XX 15-MAY-2002 (first entry)
XX
XX
DE Ovary cancer related gene sequence SFG ID NO:6207.
XX
XX
KW Human; Cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cystostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 18-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235843P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX


```

AL17037
ID AAL17037 standard; cDNA; 333 BP.
XX
AC AAL17037:
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 9494.
XX
KM Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PE 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 1697; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 333 BP; 88 A; 93 C; 79 G; 73 T; 0 other;
XX
Query Match 61.3%; Score 291.8; DB 22; Length 333;
Best Local Similarity 99.3%; Pred. No. 8.8e-80;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 CTGCACGACGACTGAACACAGACAGAGCCGCTCGCCATGAAGCTGCTGATGTCCT 65
DB 32 CTGCACGACGACTGAACACAGACAGAGCCGCTCGCCATGAAGCTGCTGATGTCCT 91
QY 66 CATGCTGGCGGCCCTCTCTCGACAGCTGATGAGATTCTGGCGCAAACTCGTGGAGGA 125
DB 92 CATGCTGGCGGCCCTCTCTCGACAGCTGATGAGATTCTGGCGCAAACTCGTGGAGGA 151
QY 126 CATGTTGAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGAGCTTCTTCA 185
DB 152 CATGGCTGAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGAGCTTCTTCA 211
QY 186 AGAGTTCTATAGACAGTGATGCGCGTGCAGAGGCTATGGGGAATTCACAGCTTTTCT 245
DB 212 AGAGTTCTATAGACAGTGATGCGCGTGCAGAGGCTATGGGGAATTCACAGCTTTTCT 271
QY 246 CAACACGTCACATGAACCTGAAAACCTTTGAGCTGATGATGATACAGCTGAC 300
DB 272 CAACACGTCACATGAACCTGAAAACCTTTGAGCTGATGATGATACAGCTGAC 326

```

```

RESULT 14
AAL08758
ID AAL08758 standard; cDNA; 370 BP.
XX
AC AAL08758:
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 1215.
XX
KM Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PE 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 270; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 370 BP; 93 A; 104 C; 94 G; 79 T; 0 other;
XX
Query Match 61.3%; Score 291.8; DB 22; Length 370;
Best Local Similarity 99.3%; Pred. No. 9.3e-80;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 CTGCACGACGACTGAACACAGACAGAGCCGCTCGCCATGAAGCTGCTGATGTCCT 65
DB 69 CTGCACGACGACTGAACACAGACAGAGCCGCTCGCCATGAAGCTGCTGATGTCCT 128
QY 126 CATGTTGAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGAGCTTCTTCA 185
DB 129 CATGCTGGCGGCCCTCTCTCGACAGCTGATGAGATTCTGGCGCAAACTCGTGGAGGA 188
QY 66 CATGCTGGCGGCCCTCTCTCGACAGCTGATGAGATTCTGGCGCAAACTCGTGGAGGA 125
DB 92 CATGCTGGCGGCCCTCTCTCGACAGCTGATGAGATTCTGGCGCAAACTCGTGGAGGA 151
QY 126 CATGTTGAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGAGCTTCTTCA 185
DB 189 CATGGCTGAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGAGCTTCTTCA 248
QY 186 AGAGTTCTATAGACAGTGATGCGCGTGCAGAGGCTATGGGGAATTCACAGCTTTTCT 245
DB 249 AGAGTTCTATAGACAGTGATGCGCGTGCAGAGGCTATGGGGAATTCACAGCTTTTCT 308
QY 246 CAACACGTCACATGAACCTGAAAACCTTTGAGCTGATGATGATACAGCTGAC 300

```


THIS PAGE BLANK (USPTO)


```
Db 1 ACAGCTGCCACGACGACTGTAACACAGACAGACGCCGCTCCGACATGAAGCTCTGATG 60
OY 61 GTCCCTATGCTGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 61 GTCCCTATGCTGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
OY 121 GAGGACATGGTGTGAAGAACCCATTAATTCGACATATCTATCTATCTGTAATACAAAGACTT 180
Db 121 GAGGACATGGTGTGAAGAACCCATTAATTCGACATATCTATCTATCTGTAATACAAAGACTT 180
OY 181 CTTTAAAGTTTCATAGAAAGTGAATGCGGCTGACAGAGCTATGGGGAATTCAGCACTGT 240
Db 181 CTTTAAAGTTTCATAGAAAGTGAATGCGGCTGACAGAGCTATGGGGAATTCAGCACTGT 240
OY 241 TTCTCAACGACGATCAGATAGAACTCTGAAAAAATTGGAGTATGATGATACATAGCTGTAC 300
Db 241 TTCTCAACGACGATCAGATAGAACTCTGAAAAAATTGGAGTATGATGATACATAGCTGTAC 300
OY 301 GACAGCATTTGGTGTAAATATGAGAGTAATTAACCTTACCAGAGGCTTTGGCTCAGAGG 360
Db 301 GACAGCATTTGGTGTAAATATGAGAGTAATTAACCTTACCAGAGGCTTTGGCTCAGAGG 360
OY 361 GCTACAGACTATGGCCAGAACTCATCTGTGATTGCTAGAGAAACCACTTCTTCTGTGT 420
Db 361 GCTACAGACTATGGCCAGAACTCATCTGTGATTGCTAGAGAAACCACTTCTTCTGTGT 420
OY 421 GCTTTTATGTGGGAACTGCTAGACAACTGTGGAACCTCAATTCATTCATTCA 476
Db 421 GCTTTTATGTGGGAACTGCTAGACAACTGTGGAACCTCAATTCATTCATTCA 476
```

RESULT 2

```
US-09-110-716-30
; Sequence 30, Application US/09110716A
; Patent No. US20020034739A1
; GENERAL INFORMATION:
; APPLICANT: Lehnert, Robert I.
; APPLICANT: Zheo, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-205961.00
; CURRENT APPLICATION NUMBER: US/09/110.716A
; CURRENT FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (41)..(325)
US-09-110-716-30
```

```
Query Match 91.2%; Score 434.2; DB 10; Length 497;
Best Local Similarity 97.8%; Pred. No. 1.2e-123;
Matches 451; Conservative 0; Mismatches 8; Indels 2; Gaps 1;
```

```
OY 6 CTGGCAGCAGAGCTGAACACAGACAGACAGCGGCTCGCCATGAAGCTGATGCTCT 65
Db 1 CTGGCAGCAGAGCTGAACACAGACAGACAGCGGCTCGCCATGAAGCTGATGCTCT 60
OY 66 CATCTGCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 125
Db 61 CATCTGCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
OY 126 CATGTTGAAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGACTTCTTCA 185
Db 121 CATGTTGAAAAAGACATCAATTCGACATATCTATACCTGAATACAAAGACTTCTTCA 180
OY 186 AGACTTCATAGACAGTATGCGGCTGACAGAGCTATGGGGAATTCAGCACTGTCTCT 245
Db 181 AGACTTCATAGACAGTATGCGGCTGACAGAGCTATGGGGAATTCAGCACTGTCTCT 240
```

```
OY 246 CAACAGTCAATAGAACTGTGAAAACTTTGACATGATGATACAGTACAGTACGACAG 305
Db 241 CAACAGTCAATAGAACTGTGAAAACTTTGACATGATGATACAGTACAGTACGACAG 300
OY 306 CATTTGTTATATATGAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 365
Db 301 CATTTGTTATATATGAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 360
OY 366 AGACTATGGCCAGAACATCTCTGTTGATGCTAGAAACCACTT -CTTCTGTGTTGCT 423
Db 361 AGACTATGGCCAGAACATCTCTGTTGATGCTAGAAACCACTTCTTCTGTGTTGCT 420
OY 424 TTTTATGTGGGAACTGCTAGACAACTGTTGAAACCTCAAT 464
Db 421 TTTTATGTGGGAACTGCTAGACAACTGTTGAAACCTCAAT 461
```

RESULT 3

```
US-09-967-768A-62
; Sequence 62, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 62
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-62
```

```
Query Match 85.6%; Score 407.4; DB 10; Length 491;
Best Local Similarity 97.0%; Pred. No. 2e-115;
Matches 426; Conservative 0; Mismatches 11; Indels 2; Gaps 1;
```

```
OY 28 GACAGCAGCGGCTCGCCAGTGAAGCTGCTGATGCTCTCTCTCTCTCTCTCTCTCTCT 87
Db 1 GACAGCAGCGGCTCGCCAGTGAAGCTGCTGATGCTCTCTCTCTCTCTCTCTCTCTCT 60
OY 88 CACTGCTATGAGATATTCGCTGCTGAACCTCTGAGGACATGTTGAAAGACATCAAT 147
Db 61 CACTGCTATGAGATATTCGCTGCTGAACCTCTGAGGACATGTTGAAAGACATCAAT 120
OY 148 TCCGACATATATATACCTGTAATACAAAGAGCTTCTTGAAGAGTTCAATAGACAGTATGCC 207
Db 121 TCCGACATATATATACCTGTAATACAAAGAGCTTCTTGAAGAGTTCAATAGACAGTATGCC 180
OY 208 GCTGACAGAGCTATGAGGAAATTAAGCAGTGTCTCTCAACAGCTACATAGAACTCTG 267
Db 181 GCTGACAGAGCTATGAGGAAATTAAGCAGTGTCTCTCAACAGCTACATAGAACTCTG 240
OY 268 AAAAATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 327
Db 241 AAAAATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
OY 328 AATTAACCTTACCAAGGCTTTGGCTCAGAGGCTACAGACTATGAGCAGAACTCATCT 387
Db 301 AATTAACCTTACCAAGGCTTTGGCTCAGAGGCTACAGACTATGAGCAGAACTCATCT 360
OY 388 GTTGATGCTAGAAACCACTT -CTTCTGTGTTGCTTTTATGTTGGAACCTGATGAC 445
Db 181 GTTGATGCTAGAAACCACTT -CTTCTGTGTTGCTTTTATGTTGGAACCTGATGAC 445
```

Db 361 GTTGATTGCTAGAAACCACTTTCTTCTGTGTGCTTTTATGAGAACTGCTAGAC 420
QY 446 AACTGTGAACCTCAAT 464
Db 421 AACTGTGAACCTCAAT 439

RESULT 4

US-09-867-701-6508
Sequence 6508, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Agilate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6508
LENGTH: 368
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-6508

Query Match 58.8%; Score 280; DB 10; Length 368;
Best Local Similarity 95.8%; Pred. No. 2,2e-76;
Matches 299; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

QY 156 ATCTATACCTAATACAAAGAGCTTCTTCAGAGTTCATACAGTGTATGCCGCTGAGA 215
Db 2 ACCTATACCTAATACAAAGAGCTTCTTCAGAGTTCATACAGTGTATGCCGCTGAGA 61
QY 216 GCGTATGGGGAATTCAGAGAGTGTTCCTCAACAGCATAGAACTGAAAACTT 275
Db 62 GCGTATGGGGAATTCAGAGAGTGTTCCTCAACAGCATAGAACTGAAAACTT 121
QY 276 TGGACTGATGATGATACAGTGTACGACAGCATTTGGTGAATATGAAGAGTAATTA 335
Db 122 TGGACTGATGATGATGATACAGTGTACGACAGCATTTGGTGAATATGAAGAGTAATTA 181
QY 336 TTACCAAGGCGTTGGCTGAGAGGCTACAGACTATGCCAGAACTCATCTGTGATG 395
Db 182 TTACCAAGGCGTTGGCTGAGAGGCTACAGACTATGCCAGAACTCATCTGTGATG 241
QY 396 CTAGAAACCACTTCTCTTCTG--TTGCTTTTATGTGGAACTGCTAGAACACTGT 452
Db 242 CTAGAAACCACTTCTCTTCTG--TTGCTTTTATGTGGAACTGCTAGAACACTGT 301
QY 453 GAAACCTCAAT 464
Db 302 GAAACCTCAAT 313

RESULT 5

US-09-956-999-5
Sequence 5, Application US/09956999
Patent No. US20020064792A1
GENERAL INFORMATION:
APPLICANT: Lincoln, Stephen
APPLICANT: Klingler, Tod M.
APPLICANT: Au-Yang, Janice
APPLICANT: Tang, Y. Tom
APPLICANT: Goold, Richard
APPLICANT: Akerblom, Ingrid E.
APPLICANT: Sellhauer, Jeffrey J.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Murry, Lynn E.
APPLICANT: Deleane, Angelo M.
APPLICANT: Levine, Wendy B.

APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
APPLICANT: Altus, Christina M.
APPLICANT: Bandman, Olga
APPLICANT: Labrie, Samuel T.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: Database for Storage and Analysis of
FILE REFERENCE: 6514-069CON
CURRENT FILING DATE: US/09/956,999
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 08/282,955
PRIOR FILING DATE: 1995-07-29
PRIOR APPLICATION NUMBER: 08/187,530
PRIOR FILING DATE: 1994-01-27
PRIOR APPLICATION NUMBER: 08/179,873
PRIOR FILING DATE: 1994-01-11
PRIOR APPLICATION NUMBER: 08/100,523
PRIOR FILING DATE: 1993-08-03
PRIOR APPLICATION NUMBER: 08/137,951
PRIOR FILING DATE: 1993-10-14
PRIOR APPLICATION NUMBER: 07/977,780
PRIOR FILING DATE: 1992-11-19
PRIOR APPLICATION NUMBER: 07/916,491
PRIOR FILING DATE: 1992-07-17
PRIOR APPLICATION NUMBER: 08/289,822
PRIOR FILING DATE: 1994-08-12
PRIOR APPLICATION NUMBER: 08/581,240
PRIOR FILING DATE: 1995-12-29
PRIOR APPLICATION NUMBER: 08/744,026
PRIOR FILING DATE: 1996-11-05
PRIOR APPLICATION NUMBER: 08/786,999
PRIOR FILING DATE: 1997-01-23
PRIOR APPLICATION NUMBER: 08/822,262
PRIOR FILING DATE: 1997-03-20
PRIOR APPLICATION NUMBER: 08/951,750
PRIOR FILING DATE: 1997-10-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 5
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
US-09-956-999-5

Query Match 47.0%; Score 223.8; DB 10; Length 495;
Best Local Similarity 71.4%; Pred. No. 4.6e-59;
Matches 325; Conservative 0; Mismatches 122; Indels 8; Gaps 2;

QY 7 TGGCAGCAGCAGTACGACAGACAGACGCGCCTGCGCATGAAAGCTGCTGATGCTC 66
Db 7 TGGCAGCAGCAGTACGACAGACAGACGCGCCTGCGCATGAAAGCTGCTGATGCTC 66
QY 67 ATGCTGGCGGCGCTCTGCTGCACTGCTATGACAGATTCGCTGCAAACTCTGAGAGC 126
Db 67 ATGCTGGCGGCGCTCTGCTGCACTGCTATGACAGATTCGCTGCAAACTCTGAGAGC 126
QY 127 ATGCTGAAAGACCATTCGACATATCTATACCTGAATACAAAGAGCTTTCTTCAA 186
Db 127 ATGCTGAAAGACCATTCGACATATCTATACCTGAATACAAAGAGCTTTCTTCAA 186
QY 187 GAGTTCATAGACAGATGATGCGCGCTGAGAGGCTATGGGAAATTCAGACAGTGTTCCTC 246
Db 187 GAGTTCATAGACAGATGATGCGCGCTGAGAGGCTATGGGAAATTCAGACAGTGTTCCTC 246
QY 247 AACCACTACATAGACTGTAAGAACTTTGAGTGTATGCTATACAGTGTAGACAGC 306
Db 247 AACCACTACATAGACTGTAAGAACTTTGAGTGTATGCTATACAGTGTAGACAGC 306
QY 307 ATTTGGTGTATATGAAGAGTAAATTAACCTTACCCAAAGCGGCTTGGCTCAGAGGCTACA 366
Db 307 AGTCTTGTGATTTT-----ATTTTAACCTTCTGCAAGACCTTTGGCTCAGAACTGCA 360

Db 436 TTTTACTACAACTACAGACAACTGTTGAACCT 470

RESULT 8

US-09-934-054-11

; Sequence 11, Application US/09934054
; Patent No. US20020107385A1

; GENERAL INFORMATION:

; APPLICANT: Akerblom, Ingrid E.

; Hillman, Jennifer L.

; Murry, Lynn E.

; Goli, Surya K.

; Hawkins, Phillip R.

; TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 2200 Sand Hill Road, Suite 100

; CITY: Menlo Park

; STATE: CA

; COUNTRY: USA

; ZIP: 94025-6936

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/934,054

; FILING DATE: 21-Aug-2001

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/747,547

; FILING DATE: <unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0077 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 503 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-09-934-054-11

Query Match 47.0%; Score 223.8; DB 10; Length 503;

Best Local Similarity 71.4%; Pred. No. 4,6e-59;

Matches 325; Conservative 0; Mismatches 122; Indels 8; Gaps 2;

Db 7 TGGCAGCAGCAGTGAACACAGACAGCGCGCTCGCCATGAAGCTCTGATGCTC 66

Db 22 TGGCAGCGCGCAGTGAACACAGCAGCGCGCTCGCCATGAAGCTCTGATGCTC 81

Db 67 ATGCTGGCGCGCT 126

Db 82 ATGCTGGCGCGCT 141

Db 127 ATGGTGAAGAGACATATTCGACATATCTACTGTAATCAAGAGCTCTTCAA 186

Db 142 GTGATTTCCAGACATATTCACAACTGCTAAGACTGAATCAAGAACTTCTTCA 201

Db 187 GAGTTCATAGACAGTATGCGCTGACAGAGCTATGCGGAATTCACAGAGTCTTCTC 246

Db 202 GAGTTCATAGACAGTATGCGCTGACAGAGCTATGCGGAATTCACAGAGTCTTCTC 261

Db 247 AACGATCATATGAACTCTGAAAACTTGGAGTATGATGATACAGTACAGACG 306

Db 262 AACCAACGATGAACTCTGAGCAATGTCGTTATGCAATTAATATGACAGC 321

Db 307 ATTTGCTATATGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 366

Db 322 AGTCTTGTGATTT-----ATTTAACTTTCTGCAAGACCTTTGGCTCAGACATGCA 375

Db 367 GACTATGGCAGCAATCATCTGTTGATGCTAGAAC--CACTTCTCTCTCTCTCT 424

Db 376 GGGTATGGTGAAGAACCACTAGAGATGCTGCAAAACACACCTCTCTCTTCTATGCT 435

Db 425 TTTATGTCGACCTGCTAGACACTGTTGAACCT 459

Db 436 TTTTACTACAACTACAGACAACTGTTGAACCT 470

RESULT 9

US-09-825-301-73

; Sequence 73, Application US/09825301

; Patent No. US20020009738A1

; GENERAL INFORMATION:

; APPLICANT: Houghton, Raymond L.

; Applicant: Dillon, David C.

; Applicant: Molesh, David A.

; Applicant: Xu, Jiangchun

; Applicant: Zehentner, Barbara

; Applicant: Persing, David H.

; TITLE OF INVENTION: METHODS, COMPOSITIONS AND KITS FOR THE DETECTION

; FILE REFERENCE: 210121.513

; CURRENT APPLICATION NUMBER: US/09/825,301

; CURRENT FILING DATE: 2001-04-02

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 73

; LENGTH: 503

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-825-301-73

Query Match 46.7%; Score 222.2; DB 10; Length 503;

Best Local Similarity 71.2%; Pred. No. 1,4e-58;

Matches 324; Conservative 0; Mismatches 123; Indels 8; Gaps 2;

Db 7 TGGCAGCAGCAGTGAACACAGACAGCGCGCTCGCCATGAAGCTCTGATGCTC 66

Db 22 TGGCAGCGCGCAGTGAACACAGCAGCGCGCTCGCCATGAAGCTCTGATGCTC 81

Db 67 ATGCTGGCGCGCT 126

Db 82 ATGCTGGCGCGCT 141

Db 127 ATGCTGAAGAGACATATTCGACATATCTACTGTAATCAAGAGCTCTTCAA 186

Db 142 GTGATTTCCAGACATATTCACAACTGCTAAGACTGAATCAAGAACTTCTTCA 201

Db 187 GAGTTCATAGACAGTATGCGCTGACAGAGCTATGCGGAATTCACAGAGTCTTCTC 246

Db 202 GAGTTCATAGACAGTATGCGCTGACAGAGCTATGCGGAATTCACAGAGTCTTCTC 261

Db 247 AACGATCATATGAACTCTGAAAACTTGGAGTATGATGATACAGTACAGACG 306

Db 262 AACCAACGATGAACTCTGAGCAATGTCGTTATGCAATTAATATGACAGC 321

Db 307 ATTTGCTATATGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 366

Db 322 AGTCTTGTGATTT-----ATTTAACTTTCTGCAAGACCTTTGGCTCAGACATGCA 375

Db 367 GACTATGGCAGCAATCATCTGTTGATGCTAGAAC--CACTTCTCTCTCTCTCT 424

Db 376 GGGTATGGTGAAGAACCACTAGAGATGCTGCAAAACACACCTCTCTTCTATGCT 435

Db 425 TTTATGTCGACCTGCTAGACACTGTTGAACCT 459


```

: APPLICANT: Wang, Tongtong
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Bennington, Angela Ann
: APPLICANT: Zehentner, Barbara
: APPLICANT: Fanger, Gary R.
: APPLICANT: Retter, Marc W.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.491C7
: CURRENT APPLICATION NUMBER: US/10/010.742
: CURRENT FILING DATE: 2001-11-30
: NUMBER OF SEQ ID NOS: 307
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 43
: LENGTH: 467
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 242..263
: OTHER INFORMATION: n = A,T,C or G
US-10-010-742-43
```

```

Query Match          41.0%; Score 195.2; DB 12; Length 467;
Best Local Similarity 69.7%; Pred. No. 2.7e-50;
Matches 294; Conservative 0; Mismatches 120; Indels 8; Gaps 2;
```

```

QY  40 CTCGCCATGAAGCTGCTGATGCTCTCATGCTGGCGCCCTCCCTCCGCACTGCTATGCA 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  467 CTCACCATGAAGTGTGCTGATGCTCTCATGCTGGCGCCCTCCCTCCGCACTGCTAGCA 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  100 GATTCGTGCTCAAAAGCTCGAGACATGTTGAAAAGACATCAATTCGACATATCT 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  407 GCTCTGCTGCTCCCTTATTGAGATGTGATTTCCAAAGACATCAATTCGACATGCT 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  160 ATACCTGAATACAAAGAGCTTCTTCAAGAGTTCATAGACAGTGGCGCTGCAGAGGCT 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  347 AAGACTGAATACAAAGAGCTTCTTCAAGAGTTCATAGACAGTGGCGCTGCAGAGGCT 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  220 ATGGGGAATTCAGAGCTGTTTCTCAACGAGTCACATAGACTCTGAAAACCTTTGGA 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  287 ATACATGAATGAAAGATGTTTNTTAAACCAACGGATGAACTMTGAGCAATGTTGAG 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  280 CTGATGATGCTACAGTGTACGACGATTTGTTAATATGAGAGATTAATTAAGTTTAC 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  227 GTGTTTATGCAATTAATATATGACGAGCTTTTGATTT-----ATTTAATCTTCT 174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  340 CCAAGGGGTTTGGCTCAGAGGCTACAGACTATGCGCAGAACTATCTGTTGATTGCTAG 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  173 GCAGAGCTTTGGCTCAGAGGCTACAGACTATGCTGAGAAACAGCAGGATTTGCTGC 114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  400 AAC--CACTTTCTTTGCTGCTCTTTTATGTGGGAACTGCTAGACACTGTTGAAC 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  113 AAACACACACTTCTTTCTTATGCTTTTATCAACAACTACAAACAAATGTTGAAC 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  458 CT 459
    ||
Db  53 CT 52
```

```

RESULT 13
US-09-757-417-46
: Sequence 46, Application US/09757417
: Patent No. US2002082216A1
: GENERAL INFORMATION:
: APPLICANT: Fanger, Gary R.
: APPLICANT: Foy, Theresa M.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Reed, Steven G.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
: FILE REFERENCE: 210121.479C1
```

```

: CURRENT APPLICATION NUMBER: US/09/757.417
: CURRENT FILING DATE: 2001-01-08
: NUMBER OF SEQ ID NOS: 49
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 46
: LENGTH: 399
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-757-417-46
```

```

Query Match          32.2%; Score 153.4; DB 10; Length 399;
Best Local Similarity 72.4%; Pred. No. 1.7e-37;
Matches 199; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
```

```

QY  46 ATGAAGCTGCTGATGCTGATGCTGATGCTGCGCCCTCTCTCGACACTGCTATGCAATCT 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  118 ATGAAGTGTGCTGATGCTGATGCTGCGCCCTCTCTCGACACTGCTATGCAAGCTCT 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  106 GCTGCAAACTCTGAGAGCATGTTGAAAAGACATCAATTCGACATATCTATACCT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  178 GCTGCCCCCTTATTGGAATGTGATTTCCAAAGACATCAATTCAGAGTGTCTAAGACT 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  166 GAATACAAAGAGCTTCTTCAAGAGTCTATAGACAGTATGCCGCTGAGAGGCTATGGG 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  238 GAATACAAAGAGCTTCTTCAAGAGTCTATAGACAGTATGCCAATTCACATATGAT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  226 AAATTCAGAGCTTCTTCTCAACCATGACATGAACTGGAATACTTTGAGCTATG 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  298 GAATTCAGAGAGATGTTTCTTAAACCAACGATGAACTCTGACCAATGTTGAGGTGTT 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  286 ATGCATACAGTGTGACAGACATTTGGTATATAT 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  358 ATGCATTAATATATGACAGAGCTTTTGATTT 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 14
US-10-007-805-492
: Sequence 492, Application US/10007805
: Patent No. US20020150581A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Yunglu
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Fanger, Gary R.
: APPLICANT: Vedick, Thomas S.
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Durham, Margarita
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.470C10
: CURRENT APPLICATION NUMBER: US/10/007.805
: CURRENT FILING DATE: 2001-12-07
: NUMBER OF SEQ ID NOS: 593
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 492
: LENGTH: 1233
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-007-805-492
```

```

Query Match          32.2%; Score 153.4; DB 12; Length 1233;
Best Local Similarity 72.4%; Pred. No. 3.1e-37;
Matches 199; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
```

```

QY  46 ATGAAGCTGCTGATGCTGATGCTGCGCCCTCTCTCGACACTGCTATGCAATCT 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1 ATGAAGTGTGCTGATGCTGATGCTGCGCCCTCTCTCGACACTGCTATGCAAGCTCT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  106 GCTGCAAACTCTGAGAGCATGTTGAAAAGACATCAATTCGACATATCTATACCT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db      61  GCGTGGCCCTATGAGAAATGATGATTCCAAGCAATCAATCCACAAAGTGTAAAGACT 120
      |||||  _  |||||  _  ||  |||||  |||||  _  |||||  ||
QY      166  GAATACAAAGAGCTTCTTCAGAGTTCATAGACAGTGTGCGGCTGCAGAGCTATGGGG 225
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db      121  GAATACAAAGAGCTTCTTCAGAGTTCATAGACAGTGTGCGGCTGCAGAGCTATGAT 180
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
QY      226  AATTCAGCAGTGTCTTCCTCAACAGTACATAGAACTGAAAACTTGGAGCTGATG 285
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db      181  GAATTGAAGGAATGTTTCTTAACCAAGCATGAAACTCTGAGCAATGTTGAGCTGTTT 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
QY      286  ATGCATACAGTGTACGACAGCATTTGGTGAATAT 320
      |||||  _  |||||  |||||  |||||  |||||  |||||  ||
Db      241  ATGCAATTAATATATGACAGCAGTCTTGTGATTT 275
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
```

RESULT 15

```
US-10-007-805-491
; Sequence 491, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, David C.
; APPLICANT: Milchem, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Finger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 491
; LENGTH: 2232
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-007-805-491
```

Query Match

32.2%: Score 153.4; DB 12; Length 2232;

Best Local Similarity 72.4%: Pred. No. 4.2e-37;

Matches 199; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
QY      46  ATGAGCTGCTGATGATGCTCATGCGGCGCCCTCCTGCACCTGCTATGACAGATTCT 105
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db      1  ATGAGTGTGCTGATGATGCTCATGCGGCGCCCTCCTGCACCTGCTATGACAGCTCT 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
QY      106  GCGTGCAAACTCTGAGAGCATGTTGAAAAGACATCAATTCGACATATATACCT 165
      |||||  _  |||||  |||||  |||||  |||||  |||||  ||
Db      61  GCGTGGCCCTATGAGAAATGATGATTCCAAGCAATCAATCCACAAGTGTAAAGACT 120
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
QY      166  GAATACAAAGAGCTTCTTCAGAGTTCATAGACAGTGTGCGGCTGCAGAGCTATGGGG 225
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db      121  GAATACAAAGAGCTTCTTCAGAGTTCATAGACAGTGTGCGGCTGCAGAGCTATGAT 180
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
QY      226  AATTCAGCAGTGTCTTCCTCAACAGTACATAGAACTGAAAACTTGGAGCTGATG 285
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db      181  GAATTGAAGGAATGTTTCTTAACCAAGCATGAAACTCTGAGCAATGTTGAGCTGTTT 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
QY      286  ATGCATACAGTGTACGACAGCATTTGGTGAATAT 320
      |||||  _  |||||  |||||  |||||  |||||  |||||  ||
Db      241  ATGCAATTAATATATGACAGCAGTCTTGTGATTT 275
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
```

Search completed: January 25, 2003, 11:40:07

Job time : 89 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 10:18:48 ; Search time 73 seconds
(without alignments)
1999.701 Million cell updates/sec

Title: US-09-806-302A-1

Perfect score: 476
Sequence: 1 acgagctgcacgcacgact.....ccatcatcattca 476

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCU08.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	476	100.0	476	3	US-08-821-451A-5
2	476	100.0	476	4	US-09-263-810-5
3	476	100.0	476	4	US-09-583-169-5
4	223.8	47.0	495	4	US-08-969-987-5
5	223.8	47.0	503	1	US-08-455-896-1
6	223.8	47.0	503	2	US-08-933-149-1
7	223.8	47.0	503	2	US-09-082-343-1
8	223.8	47.0	503	3	US-09-082-253-1
9	223.8	47.0	503	5	PCT-US96-08235-1
10	223.8	47.0	535	4	US-09-215-818-1
11	201.2	42.3	403	1	US-08-455-896-5
12	201.2	42.3	403	2	US-08-933-149-5
13	201.2	42.3	403	2	US-09-082-343-5
14	201.2	42.3	403	3	US-09-082-253-5
15	201.2	42.3	403	5	PCT-US96-08235-5
16	57.2	12.0	206	1	US-08-455-896-6
17	57.2	12.0	206	2	US-08-933-149-6
18	57.2	12.0	206	2	US-09-082-343-6
19	57.2	12.0	206	2	US-09-082-253-6
20	57.2	12.0	206	5	PCT-US96-08235-6
21	33	6.9	2685	4	US-09-061-768A-1
22	32.6	6.8	35524	3	US-08-923-137-1
23	31.4	6.6	1528	1	US-08-459-586-12
24	31.4	6.6	1528	2	US-08-282-696-12
25	31.4	6.6	2047	4	US-08-836-261A-1
26	31.4	6.6	2489	1	US-08-459-586-1
27	31.4	6.6	2489	2	US-08-282-696-1

C 28	31.4	6.6	7242	1	US-08-459-586-3	Sequence 3, Appl1
C 29	31.4	6.6	7242	2	US-08-282-696-3	Sequence 3, Appl1
C 30	31.4	6.6	18994	1	US-08-459-586-4	Sequence 4, Appl1
C 31	31.4	6.6	18994	2	US-08-282-696-4	Sequence 4, Appl1
C 32	31	6.5	2842	1	US-08-175-388-3	Sequence 3, Appl1
C 33	31	6.5	2842	2	US-08-779-620-3	Sequence 3, Appl1
C 34	31	6.5	2842	2	US-08-818-726-3	Sequence 3, Appl1
C 35	31	6.5	2889	1	US-08-537-002A-4	Sequence 4, Appl1
C 36	31	6.5	2889	3	US-08-863-010-4	Sequence 4, Appl1
C 37	31	6.5	2889	4	US-08-537-002A-5	Sequence 4, Appl1
C 38	31	6.5	3600	1	US-08-537-002A-5	Sequence 5, Appl1
C 39	31	6.5	3600	3	US-08-863-010-5	Sequence 5, Appl1
C 40	31	6.5	3600	3	US-09-024-429-5	Sequence 5, Appl1
C 41	31	6.5	152331	3	US-09-128-155-16	Sequence 16, Appl1
C 42	30.8	6.5	4027	1	US-08-348-143-3	Sequence 3, Appl1
C 43	30.8	6.5	4027	1	US-08-348-143-4	Sequence 4, Appl1
C 44	30.8	6.5	4027	1	US-08-571-785-3	Sequence 3, Appl1
C 45	30.8	6.5	4027	1	US-08-571-785-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-821-451A-5
; Sequence 5, Application US/08821451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996 ,
; ATTORNEY/AGENT INFORMATION:
; NAME: MULHINS, J.C.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEO ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; US-08-821-451A-5
Query Match 100.0%; Score 476; DB 3; Length 476;
Best Local Similarity 100.0%; Pred. No. 6.8e-142; Indels 0; Gaps 0;
Matches 476; Conservative 0; Mismatches 0;
QY 1 ACGAGCTGCACGACGACTGACACAGACAGCCGCTCGCCATGAGCTGCTGATG 60

```
Db 1 ACGAGCTGCCACGACGACTGACACAGACAGCCGCTCGCATGAACTGCTGATG 60
Oy 61 GTCCCTCATGCTGGGGGCGCTCTCTGCACTGCAATGAGATTTGGCTGCAACTCTG 120
Db 61 GTCCCTCATGCTGGGGGCGCTCTCTGCACTGCAATGAGATTTGGCTGCAACTCTG 120
Oy 121 GAGGACATGCTTGAAGAACCATCAATTCGACATATCTATCTAATACAAAGACTT 180
Db 121 GAGGACATGCTTGAAGAACCATCAATTCGACATATCTAATACAAAGACTT 180
Oy 181 CTTCAAGAGTTTCATAGACAGTATGCGCTGACAGAGCTATGGGAAATTCAGACAGT 240
Db 181 CTTCAAGAGTTTCATAGACAGTATGCGCTGACAGAGCTATGGGAAATTCAGACAGT 240
Oy 241 TTCTCAACGAGTCATAGACAGTATGCGCTGAGAACTTTGAGCTGATGATGATGATG 300
Db 241 TTCTCAACGAGTCATAGACAGTATGCGCTGAGAACTTTGAGCTGATGATGATGATG 300
Oy 301 GACAGCATTTGGTATATGATGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Db 301 GACAGCATTTGGTATATGATGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Oy 361 GCTACAGACTTGGCCAGCACTCATCTGTTGATGCTAGAAACACATTTCTTCTGTT 420
Db 361 GCTACAGACTTGGCCAGCACTCATCTGTTGATGCTAGAAACACATTTCTTCTGTT 420
Oy 421 GCTTTTATGTGGGAACTGCTAGACAACTGTTGAACCTCAATTCATTCATTTCA 476
Db 421 GCTTTTATGTGGGAACTGCTAGACAACTGTTGAACCTCAATTCATTCATTTCA 476
```

RESULT 2

```
US-09-263-810-5
: Sequence 5, Application US/09263810
: Patent No. 6174992
: GENERAL INFORMATION:
: APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
: TITLE OF INVENTION: Human Endometrial Specific Steroid-
: TITLE OF INVENTION: Binding Factor I, II and III
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
: ADDRESSEE: CECCHI, STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/263,810
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/821,451
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: MULLINS, J.G.
: REGISTRATION NUMBER: 33,073
: REFERENCE/DOCKET NUMBER: 325800-521 (P257)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1740
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 476 BASE PAIRS
: TYPE: NUCLEIC ACID
: STRANDEDNESS: SINGLE
```

```
: TOPOLOGY: LINEAR
: MOLECULE TYPE: cDNA
US-09-263-810-5
```

```
Query Match 100.0%; Score 476; DB 4; Length 476;
Best Local Similarly 100.0%; Pred. No. 6,8e-142;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 ACGAGCTGCCACGACGACTGACACAGACAGCCGCTCGCATGAACTGCTGATG 60
Db 1 ACGAGCTGCCACGACGACTGACACAGACAGCCGCTCGCATGAACTGCTGATG 60
Oy 61 GTCCCTCATGCTGGGGGCGCTCTCTGCACTGCAATGAGATTTGGCTGCAACTCTG 120
Db 61 GTCCCTCATGCTGGGGGCGCTCTCTGCACTGCAATGAGATTTGGCTGCAACTCTG 120
Oy 121 GAGGACATGCTTGAAGAACCATCAATTCGACATATCTAATACAAAGACTT 180
Db 121 GAGGACATGCTTGAAGAACCATCAATTCGACATATCTAATACAAAGACTT 180
Oy 181 CTTCAAGAGTTTCATAGACAGTATGCGCTGACAGAGCTATGGGAAATTCAGACAGT 240
Db 181 CTTCAAGAGTTTCATAGACAGTATGCGCTGACAGAGCTATGGGAAATTCAGACAGT 240
Oy 241 TTCTCAACGAGTCATAGACAGTATGCGCTGAGAACTTTGAGCTGATGATGATGATG 300
Db 241 TTCTCAACGAGTCATAGACAGTATGCGCTGAGAACTTTGAGCTGATGATGATGATG 300
Oy 301 GACAGCATTTGGTATATGATGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Db 301 GACAGCATTTGGTATATGATGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Oy 361 GCTACAGACTTGGCCAGCACTCATCTGTTGATGCTAGAAACACATTTCTTCTGTT 420
Db 361 GCTACAGACTTGGCCAGCACTCATCTGTTGATGCTAGAAACACATTTCTTCTGTT 420
Oy 421 GCTTTTATGTGGGAACTGCTAGACAACTGTTGAACCTCAATTCATTCATTTCA 476
Db 421 GCTTTTATGTGGGAACTGCTAGACAACTGTTGAACCTCAATTCATTCATTTCA 476
```

RESULT 3

```
US-09-583-169-5
: Sequence 5, Application US/09583169
: Patent No. 6338948
: GENERAL INFORMATION:
: APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz
: TITLE OF INVENTION: Human Endometrial Specific Steroid-
: TITLE OF INVENTION: Binding Factor I, II and III
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
: ADDRESSEE: CECCHI, STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/583,169
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/821,451
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: MULLINS, J.G.
: REGISTRATION NUMBER: 33,073
```

```

: REFERENCE/DOCKET NUMBER: 325800-521 (Pf257)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 476 BASE PAIRS
: TYPE: NUCLEIC ACID
: STRANDEDNESS: SINGLE
: TOPOLOGY: LINEAR
: MOLECULE TYPE: CDNA
: US-09-583-169-5

Query Match      100.0%; Score 476; DB 4; Length 476;
Best Local Similarity 100.0%; Pred. No. 6.8e-142;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      1  ACGAGCTGCCACGACGACGACTGAAACACACAGACAGACCCGCTGCCATGAAGCTGCGTGAAG 60
Db      1  ACGAGCTGCCACGACGACGACTGAAACACACAGACAGACCCGCTGCCATGAAGCTGCGTGAAG 60

QY      61  GTCCCTCATGCTGGGGGGCCCTCCTCTCGACACGTCATGACAGATTCTGCGTCGCAAACTCCG 120
Db      61  GTCCCTCATGCTGGGGGGCCCTCCTCTCGACACGTCATGACAGATTCTGCGTCGCAAACTCCG 120

QY      121  GAGGACATGTTTGAAGAAACATCATTTCCGACATATCTATACCTGGAATACAAAGACTT 180
Db      121  GAGGACATGTTTGAAGAAACATCATTTCCGACATATCTATACCTGGAATACAAAGACTT 180

QY      181  CTTCAAGAGTTCATGACAGATGATCCGCTGACAGAGGCGTATGGGGAAATTCAGACAGTCT 240
Db      181  CTTCAAGAGTTCATGACAGATGATCCGCTGACAGAGGCGTATGGGGAAATTCAGACAGTCT 240

QY      241  TTCCCTCAACCAAGTCACATAGAACTCTGAAAAAACTTTGACTGATGATGCATACAGTGTAC 300
Db      241  TTCCCTCAACCAAGTCACATAGAACTCTGAAAAAACTTTGACTGATGATGCATACAGTGTAC 300

QY      301  GACAGCATTTGCTGTAATATGAAAGTAAATTACTTTTACCAGGCGTTTGCTCAGAGC 360
Db      301  GACAGCATTTGCTGTAATATGAAAGTAAATTACTTTTACCAGGCGTTTGCTCAGAGC 360

QY      361  GCTACAGACTTATGGCGCAACATCATCTGTATGCTGTGAAGCAACTTCTCTGTGTGT 420
Db      361  GCTACAGACTTATGGCGCAACATCATCTGTATGCTGTGAAGCAACTTCTCTGTGTGT 420

QY      421  GCTTTTATGTGGGAACCTGCTAGACAACTGTGTGAACCTCAATTATTCATTCA 476
Db      421  GCTTTTATGTGGGAACCTGCTAGACAACTGTGTGAACCTCAATTATTCATTCA 476

RESULT 4
US-08-969-987-5
: Sequence 5, Application us/08969987A
: Patent No. 6303297
: GENERAL INFORMATION:
: APPLICANT: Lincoln, Steve
: APPLICANT: Klinger, Tod M.
: APPLICANT: Au-Young, Janice
: APPLICANT: Tang, Y. Tom
: APPLICANT: Goold, Richard
: APPLICANT: Akerblom, Ingrid E.
: APPLICANT: Seilheimer, Jeffrey J.
: APPLICANT: Hawkins, Phillip R.
: APPLICANT: Murry, Lynn E.
: APPLICANT: Deleage, Angelo M.
: APPLICANT: Levine, Wendy B.
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Goli, Surya K.
: APPLICANT: Altus, Christina M.
: APPLICANT: Bandman, Olga
: APPLICANT: Labrie, Samuel T.
: APPLICANT: Shah, Puryl
: TITLE OF INVENTION: Database for Storage and Analysis of

```

```

TITLE OF INVENTION: Full Length Sequences
FILE REFERENCE: 6514-069001
CURRENT APPLICATION NUMBER: US/08/969, 987A
CURRENT FILING DATE: 1997-11-13
EARLIER APPLICATION NUMBER: 08/282, 955
EARLIER FILING DATE: 1995-07-29
EARLIER APPLICATION NUMBER: 08/187, 530
EARLIER FILING DATE: 1994-01-27
EARLIER APPLICATION NUMBER: 08/179, 873
EARLIER FILING DATE: 1994-01-11
EARLIER APPLICATION NUMBER: 08/100, 523
EARLIER FILING DATE: 1993-08-03
EARLIER APPLICATION NUMBER: 07/977, 780
EARLIER FILING DATE: 1992-11-19
EARLIER APPLICATION NUMBER: 07/916, 491
EARLIER FILING DATE: 1992-07-17
EARLIER APPLICATION NUMBER: 08/289, 822
EARLIER FILING DATE: 1994-06-12
EARLIER APPLICATION NUMBER: 08/581, 240
EARLIER FILING DATE: 1995-12-29
EARLIER APPLICATION NUMBER: 08/657, 697
EARLIER FILING DATE: 1996-05-29
EARLIER APPLICATION NUMBER: 08/747, 547
EARLIER FILING DATE: 1996-11-12
EARLIER APPLICATION NUMBER: 08/712, 710
EARLIER FILING DATE: 1996-09-12
EARLIER APPLICATION NUMBER: 08/744, 026
EARLIER FILING DATE: 1996-11-05
EARLIER APPLICATION NUMBER: 08/786, 999
EARLIER FILING DATE: 1997-01-23
EARLIER APPLICATION NUMBER: 08/822, 262
EARLIER FILING DATE: 1997-03-20
EARLIER APPLICATION NUMBER: 08/951, 750
EARLIER FILING DATE: 1997-10-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
US-08-969-987-5

Query Match      47.0%; Score 223.8; DB 4; Length 495;
Best Local Similarity 71.4%; Pred. No. 1,1e-61;
Matches 325; Conservative 0; Mismatches 122; Indels 8; Gaps 2;

QY      7      TGGCAGCGACGACGCTAAACACAGACAGCAGCCGCCCTGCCATGAAGCTGCTGATGCTCTC 66
Db      7      TGGCAGCCGCGACGCTGAACACCGACGACGACGACACCTCCACCATGAACTGGTATGGTGCCTC 66
QY      67      ATGCGTGGGGGCGCCCTCTCTGTCATGCTATGAGATTCGAGTTCGGCTGGCAACTCTTGAGGAGC 126
Db      67      ATGCTGGGGGCGCCCTCTCCAGCAGCTGCTTAGCGAGGCTGTGGCTGCCCTTATTCGAGAAAT 126
QY      127     ATGTTGGAAGAAGACCATGATTCGAGCATATCTATACCTGAATACGAAGAAGCTTCTTCAA 186
Db      127     GTGATTTCCAGACAAATCATTCACAGAAGTGCTTAGAGACTGAATACGAAGAAGACTTCTTCAA 186
QY      187     GAGTTCATAGACAGCTGATGCGCGCTGCAGAGGCTATGGGGAATTTCAAGCAGTGTTTCTC 246
Db      187     GAGTTCATATGACAGCAATTCGCCATTCGAATGAGATTTGAAGGAATGTTTCTT 246
QY      247     AACCAAGTACATAGAACTCTGAAAAACTTTGGAGCTGATGATGATCATAGTATGACAGAC 306
Db      247     AACCAAGAGATGAAATCTGTGAGCAATGTTGAGGTGTTATTCATTAATATATGACAGC 306
QY      307     ATTGGGTGTAATATGACAGATTAATTAACTTTACCCGAAGCGGTTTGGCTCAGAGGCGCTACA 366
Db      307     AGCTTTTGTGATTT-----ATTTAACCTTCTGCAAGACGCTTTGGCTCAGAGACTGCA 360
QY      367     GACTATGGCCAGCAACTCATCTGTATTGATTCGTAAGAAC--CACTTTCCTTGTGTGCTT 424
Db      361     GGGTATGGTGGAGAACCAACTACGATGATTCGCAACACCAACACACTTCTCTTTCTTATGTCT 420

```


OY		187	GAGTTCATGACCATGTGGCCCGCTGCAGAGCCTATGGCGAAATTCAAGAAGTGTTTCCTC	246
Db		202	GAGTTCATGACGACGAATGCCACTACAATAATGCCATAGATGTAATTGANGAATGTTTTCTT	261
OY		247	AACCAGTCACATAGAACCTGTGAAAAAAGCTTGGACTGTATGATGCATACAGTGTACGAAGC	306
Db		262	AACCAAACGGATGAAACTCTGACCAATGTGAGGTGTTTATGCATAATATATATGACAGC	321
OY		307	ATTGGGTATATATGAAAGTAGTAATTAATCACTTAAOCCAAGGCGTTGGCTCAGAGCGCTACA	366
Db		322	AGCTCTGTGTGATTT-----ATTTAACCTTTCGCAAGACCTTGGCTCACGAACACTGCA	375
OY		367	GACATATGCCCGAACTCATCTGTGTGATGGCTAGAAAAC-CACATTCTCTCTTGTTGGCTT	424
Db		376	GGGATATGGTSGAAGAACCAACTACGATATGGTCGCAAAACACACACCCCTCTCTTTATATGCT	435
OY		425	TTTATGTGGGAACGTGTAGAGCAACTGTGTGNAACCT	459
Db		436	TTTTACTTCAACATACAGAAATTTGTGAAACCT	470

```

US-09-082-343-1
: Sequence 1, Application US/09082343
: Patent No. 5968754
: GENERAL INFORMATION:
: APPLICANT: WATSON, MARK A.
: APPLICANT: FLEMING, TIMOTHY P.
: TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
: TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
: STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
: CITY: ST. LOUIS
: STATE: MISSOURI
: COUNTRY: USA
: ZIP: 63105-1817
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/082,343
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/455,896
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: HOLLAND, DONALD R.
: REGISTRATION NUMBER: 35,197
: REFERENCE/DOCKET NUMBER: 952726
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (314) 727-5188
: TELEFAX: (314) 727-6092
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 503 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-09-082-343-1

```

Query Match	47.0%;	Score 223.8;	DB 2;	Length 503;
Best Local Similarity	-71.4%;	Pred. No. 1.1e-61;		
Matches 325; Conservative	0;	Mismatches 122;	Indels 8;	Gaps 2;

OY	7	TGCCACGACGACGTAAGACACAGACAGAGAGGAGGCGGCTCGCCATTTGAAAGGTGGTGATGGTCC	66
Db	22	TGCCACGCGGACACTBAACACCGACGACGAGCAGCGCTCACCATGAGATTGCTGATGGTCC	81
OY	67	ATGCTGGGCGCCCTCCCTCCGCTCAGCTGGTATGACATTTCTGGCTGCAAACTCCCTGGAGGAC	126
Db	82	ATGCTGGGCGCCCTCCCTCCGCTCAGCTGGTACGAGGCTCTGGCTGCGCCCTTATGGAGAAAT	141
OY	127	ATGGTTGAAAAGACCACTTCAATTCGCAATATCTATACCTGATTCAAAAGGCTTTCCAA	186
Db	142	GTGATTTTCCAAAGAACATCAATCCCAAGTGTCTAAAGCTAATTCAAAAGAACTTCTTCAA	201
OY	187	GAGTTTCATAGACAGTGTATGCGCGTGCAGAGGCTATGGGAAATTCAGACGCTGTTCCCTC	246
Db	202	GAGTTTCATAGACAGCAATGCCACTACCAATGCCATTAATGATTAATGAAAGAAATGTTTCTT	261
OY	247	AACCACTCACATAGNACTCTGAAAACCTTTGGACTGATGATGCATACAGTGCAGCAGC	306
Db	262	AACCAAAAGGATGGAACACTCTGAGCAATGTGGAGGTGTTTATGCAATTAATATGACAGC	321
OY	307	ATTGGTGTATATGAGAGTAATTTAACTTTAACCCAGAGGGTTGGCTCAGAGGGCTACA	366
Db	322	AGCTTTTGTGATTT-----ATTTTAACTTTCTCAAGACCTTTGGCTCACAGACTGCA	375
OY	367	GACTATGGCGAAACTCATCTGTTGATTTGCTGTAAGAC--CACTTTCTTCTTGTGTTGCTT	424
Db	376	GGGTATGGTGAAGAAACCAACTACGGAATGGCTGCAAAACACACGCTTCTCTTTATATGCT	435
OY	425	TTTATGTGGGAACCTGTAGAACACTGTTGAAACCT	459
Db	436	TTTTACTACAACACTACGAACAATTTGTTGAAACT	470

```

1      RESULT 8
2      US-09-082-253-1
3      : Sequence 1, Application US/09082253
4      : Patent No. 6004756
5      : GENERAL INFORMATION:
6      : APPLICANT: WATSON, MARK A.
7      : APPLICANT: FLEMING, TIMOTHY P.
8      : TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
9      : TITLE OF INVENTION: MAAMARY-SPECIFIC BREAST CANCER PROTEIN
10     : NUMBER OF SEQUENCES: 13
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
13     : STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
14     : CITY: ST. LOUIS
15     : STATE: MISSOURI
16     : COUNTRY: USA
17     : ZIP: 63105-1817
18     :
19     : COMPUTER READABLE FORM:
20     : MEDIUM TYPE: Floppy disk
21     : COMPUTER: IBM PC compatible
22     : OPERATING SYSTEM: PC-DOS/MS-DOS
23     : SOFTWARE: PatentIn Release #1.0, Version #1.25
24     :
25     : CURRENT APPLICATION DATA:
26     : APPLICATION NUMBER: US/09/082,253
27     : FILING DATE:
28     : CLASSIFICATION:
29     : PRIOR APPLICATION DATA:
30     : APPLICATION NUMBER: 08/455,896
31     : FILING DATE: 05/31/1995
32     :
33     : ATTORNEY/AGENT INFORMATION:
34     : NAME: HOLLAND, DONALD R.
35     : REGISTRATION NUMBER: 35,197
36     : REFERENCE/DOCKET NUMBER: 952726
37     :
38     : TELECOMMUNICATION INFORMATION:
39     : TELEPHONE: (314) 727-5188
40     : TELEFAX: (314) 727-6092
41     :
42     : INFORMATION FOR SEQ ID NO: 1:
43     : SEQUENCE CHARACTERISTICS:
44     : LENGTH: 503 base pairs
45     : type: nucleic acid

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 base pairs
; TYPE: nucleic acid

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-082-253-1

Query Match 47.0%; Score 223.8; DB 3; Length 503;
Best Local Similarity 71.4%; Pred. No. 1,1e-61;
Matches 325; Conservative 0; Mismatches 122; Indels 8; Gaps 2;

QY 7 TGCACGCGACGCTGTAACACAGACAGAGCGGCTCCGCAATGAAGCTGATGTCCTC 66
DB 22 TGCACCGCGACGCTGTAACACAGAGCGGCTCCGCAATGAAGCTGATGTCCTC 81
DB 82 ATGCTGGCGGCGCTCTCCGCACTGCTACGAGGCTCTGCTCCCTTAATGAGAAAT 141
QY 127 ATGCTGAAAGACCATCATTCGACATATCTATACCTGAATACAAAGACTCTTCAA 186
DB 142 GTGATTTCCAGACATCAATCCACAAAGTGTCTAAGACTGATACAAAGACTCTTCAA 201
QY 187 GAGTTATAGACAGTATGCGGCTGCAGAGGCTATGGGAAATTCAGCAGTGTTCCTC 246
DB 202 GAGTTATAGACAGCAATGCGCACTACAAATGCAATGATGATGAAAGATGTTCTT 261
QY 247 AACGATGACATAGACATCTGAAAACTTGAGTATGATGATGATGATGATGATGATG 306
DB 262 AACCAAGGATGAACTGAGCAATGTTAGTGTATGCAATTAATTAATTAATTAATTA 321
QY 307 ATTTGCTGTAATATGAGAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 366
DB 322 AGCTTTGTGATTT-----ATTTAACTTCTGCAAGACCTTTGGCTCACAGAACTGCA 375
QY 367 GACTATGCGGCAACTATCTGTTGATGTTGCTAGAAAC--CACTTCTCTCTGTTGCTT 424
DB 376 GGGATGTTGTAAGAAACCACTACGATGCTGCAAAACCACTCTCTCTTCTTATGTC 435
QY 425 TTTATGCGAAGTCTAGACACTGTTGAAACCT 459
DB 436 TTTTACTACAACTACAGACAAATGTTGAAACCT 470

RESULT 9
PCT-US96-08235-1

Sequence 1, Application PC/TUS9608235
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08235
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 964796

TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 base pairs
type: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US96-08235-1

Query Match 47.0%; Score 223.8; DB 5; Length 503;
Best Local Similarity 71.4%; Pred. No. 1,1e-61;
Matches 325; Conservative 0; Mismatches 122; Indels 8; Gaps 2;

QY 7 TGCACGCGACGCTGTAACACAGACAGAGCGGCTCCGCAATGAAGCTGATGTCCTC 66
DB 22 TGCACCGCGACGCTGTAACACAGAGCGGCTCCGCAATGAAGCTGATGTCCTC 81
QY 67 ATGCTGGCGGCGCTCTCCGCACTGCTACGAGATTCGCTGCAAACTCTGAGAGAC 126
DB 82 ATGCTGGCGGCGCTCTCCGCACTGCTACGAGGCTCTGCTCCCTTAATGAGAAAT 141
QY 127 ATGCTGAAAGACCATCATTCGACATATCTATACCTGAATACAAAGACTCTTCAA 186
DB 142 GTGATTTCCAGACATCAATCCACAAAGTGTCTAAGACTGATACAAAGACTCTTCAA 201
QY 187 GACTATGACAGTATGCGGCTGCAGAGGCTATGGGAAATTCAGCAGTGTTCCTC 246
DB 202 GAGTTATAGACAGCAATGCGCACTACAAATGCAATGATGATGAAAGATGTTCTT 261
QY 247 AACGATGACATAGACATCTGAAAACTTGAGTATGATGATGATGATGATGATGATG 306
DB 262 AACCAAGGATGAACTGAGCAATGTTAGTGTATGCAATTAATTAATTAATTAATTA 321
QY 307 ATTTGCTGTAATATGAGAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 366
DB 322 AGCTTTGTGATTT-----ATTTAACTTCTGCAAGACCTTTGGCTCACAGAACTGCA 375
QY 367 GACTATGCGGCAACTATCTGTTGATGTTGCTAGAAAC--CACTTCTCTCTGTTGCTT 424
DB 376 GGGATGTTGTAAGAAACCACTACGATGCTGCAAAACCACTCTCTCTTCTTATGTC 435
QY 425 TTTATGCGAAGTCTAGACACTGTTGAAACCT 459
DB 436 TTTTACTACAACTACAGACAAATGTTGAAACCT 470

RESULT 10
US-09-215-818-1

Sequence 1, Application US/09215818A
Patent No. 6379671
GENERAL INFORMATION:
APPLICANT: Colidites, Tracey
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
TITLE OF INVENTION: DETECTING DISEASES OF THE BREAST
FILE REFERENCE: 5972-US-P2
CURRENT APPLICATION NUMBER: US/09/215,818A
CURRENT FILING DATE: 1998-12-18
EARLIER APPLICATION NUMBER: 08/912,276
EARLIER FILING DATE: 1997-08-17
EARLIER APPLICATION NUMBER: 08/697,105
EARLIER FILING DATE: 1996-08-19
EARLIER APPLICATION NUMBER: 08/912,149
EARLIER FILING DATE: 1997-08-15
EARLIER APPLICATION NUMBER: 08/697,106
EARLIER FILING DATE: 1996-08-19
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 11:33:34 : Search time 2256 Seconds
(without alignments)
3417.133 Million cell updates/sec

Title: US-09-806-302a-1

Perfect score: 476

Sequence: 1 acgagctgcacacacact.....cctcaatcattcattca 476

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estnu:*
6: em_estnu:*
7: em_estnu:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	429.2	90.2	636	10	AW966513 EST378587
2	429.2	90.2	647	10	AW966509 EST378583
3	408.2	85.8	464	10	AW451131 UT-R-B13
4	407.4	85.6	491	9	AA393164
5	405.2	85.1	434	13	BI964837
6	397.4	83.5	464	9	AI800231

C 7	391.6	82.3	460	9	AI659370	AI659370 tui1c09.x
C 8	389.6	81.8	471	9	AI936084	AI936084 w61d06.x
C 9	389	81.7	458	10	AW779377	AW779377 hn79g08.x
C 10	385.2	80.9	458	9	AI491987	AI491987 t007f11.x
C 11	378.2	79.5	366	9	AA297402	AA297402 EST112936
C 12	366.6	77.0	435	9	AA398560	AA398560 zc73f02.s
C 13	364.6	76.6	435	9	AA493295	AA493295 n976e06.s
C 14	357	75.5	365	10	BE044895	BE044895 hn06d03.x
C 15	343	72.1	351	10	BE044893	BE044893 hn06d04.x
C 16	341.2	71.7	418	10	AW243774	AW243774 x056b10.x
C 17	338	71.0	337	10	AW510561	AW510561 x037g03.x
C 18	335.4	70.5	456	10	AW966498	AW966498 EST378572
C 19	294	61.8	378	13	BI963186	BI963186 ie57c11.x
C 20	280.2	58.9	318	9	AA298750	AA298750 EST114585
C 21	280	58.8	368	9	AA525178	AA525178 n151c09.s
C 22	249.6	52.4	258	9	AA297452	AA297452 EST112981
C 23	247	51.9	253	9	AA297456	AA297456 EST112985
C 24	240	50.4	242	9	AA297405	AA297405 EST112939
C 25	206.4	43.4	463	9	AI791322	AI791322 ae54e06.y
C 26	191.8	40.3	510	10	AW972474	AW972474 EST384565
C 27	190.2	40.0	388	10	AW175618	AW175618 RC3-BT004
C 28	188	39.5	447	9	AA513640	AA513640 nh79e11.s
C 29	186.8	39.2	372	9	AI904282	AI904282 PM-BT046-
C 30	186.2	39.1	375	9	AI904278	AI904278 PM-BT046-
C 31	185	38.9	189	9	AA297482	AA297482 EST113009
C 32	184.8	38.8	444	9	AA991451	AA991451 os52f11.s
C 33	181.4	38.1	446	9	AI732534	AI732534 nh79e11.x
C 34	169	35.5	421	9	AI124690	AI124690 am60f04.x
C 35	165.4	34.7	184	9	AI937042	AI937042 wP72d06.x
C 36	160	33.6	197	12	BF745480	BF745480 CM2-BT082
C 37	157	32.0	309	14	BQ304242	BQ304242 RC1-BT081
C 38	155.2	32.6	294	9	AI905273	AI905273 RC1-BT081
C 39	143	30.0	263	9	AI905265	AI905265 RC-BT081
C 40	140.6	29.5	332	9	AA513647	AA513647 nh79f09.s
C 41	113	23.7	282	10	AW175603	AW175603 RC3-BT004
C 42	112.2	23.6	227	12	BG219371	BG219371 RST9128
C 43	110.6	23.2	193	10	AW207696	AW207696 UI-R-B12-
C 44	104.4	21.9	511	13	BI278959	BI278959 UI-R-DB0-
C 45	104	21.8	510	13	BI285288	BI285288 UI-R-DB0-

ALIGNMENTS

RESULT 1
LOCUS AW966513 636 bp mRNA linear EST 01-JUN-2000
DEFINITION EST378587 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
ACCESSION AW966513
VERSION AW966513.1 GI:8156349
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 636)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 229
Seq primer: Forward.
Location/Qualifiers
1. 636
/organism="Homo sapiens"

FEATURES

source

[illegible][illegible]

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/db/brp/image/image.html

Insert Length: 575 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 453

FEATURES

source

```
1. .460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2137932"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
150552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
```

BASE COUNT 131 a 97 c 105 g 131 t

ORIGIN

Query Match 83.5%; Score 397.4; DB 9; Length 464;

Best Local Similarity 96.8%; Pred. No. 4.5e-103;

Matches 417; Conservative 0; Mismatches 11; Indels 3; Gaps 1;

```

37 CGCCTCGCCATGAAGCTGTGATGTCCTCATGCTGGCGCCCTCCTCTGCACTGCTAT 96
|||||
464 CGCCTCGCCATGAAGCTGTGATGTCCTCATGCTGGCGCCCTCCTCTGCACTGCTAT 405
|||||
97 GCAGATTCCTGCTGCAACTCTCTGGAGACATGCTTGAAGAACCATCAATTCGGACATA 156
|||||
404 GCAGATTCCTGCTGCAACTCTCTGGAGACATGCTTGAAGAACCATCAATTCGGACATA 345
|||||
157 TCATATCCTGAATAGAAAGAGCTTCTTCAAGAGTTCATAGACAGTGTGCGCGTGCAGAG 216
|||||
344 TCATATCCTGAATAGAAAGAGCTTCTTCAAGAGTTCATAGACAGTGTGCGCGTGCAGAG 285
|||||
217 GCTATGGGGAATTCAGACAGTGTTCCTCAACAGCTACATAGAACTGGAAGAACTTT 276
|||||
284 GCTATGGGGAATTCAGACAGTGTTCCTCAACAGCTACATAGAACTGGAAGAACTTT 225
|||||
277 GCACTGATGATGATACAGTGTACAGACATTTGGTGTATATGAAGATTAATTAATT 336
|||||
224 GCACTGATGATGATACAGTGTACAGACATTTGGTGTATATGAAGATTAATTAATT 165
|||||
337 TACCAAGGCGTTGGCTACAGAGGCTACAGACTATGGCCAGAACTCATGTTGATTCG 396
|||||
164 TACCAAGGCGTTGGCTACAGAGGCTACAGACTATGGCCAGAACTCATGTTGATTCG 105
|||||
397 TAGAAACCACTTCTTCTTGTG--TTGCTTTTATGTGGAACTGCTAGACAACGTGG 453
|||||
104 TAGAAACCACTTCTTCTTGTGTTCTTTTATGTGGAACCTGCTAGACAACGTGG 45
|||||
454 AAACCTCAATT 464
|||||
44 AAACCTCAAAAT 34
```

RESULT 7

LOCUS

AI659370 460 bp mRNA linear EST 15-DEC-1999

DEFINITION

similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR.; mRNA

sequence.

AI659370

VERSION

AI659370.1 GI:4762940

KEYWORDS

EST.

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 460) NCI_CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgap@rs-remail.nih.gov Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/db/brp/image/image.html Insert Length: 559 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 448.

FEATURES

source

```
1. .460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2250736"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 132 a 94 c 101 g 133 t

ORIGIN

Query Match 82.3%; Score 391.6; DB 9; Length 460;

Best Local Similarity 97.2%; Pred. No. 2e-101;

Matches 410; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

```

46 ATGAGCTGCTGATGCTGCTACAGTGGGCGCCCTCCTCTGCACTGCTATGCAATTC 105
|||||
460 ATGAGCTGCTGATGCTGCTACAGTGGGCGCCCTCCTCTGCACTGCTATGCAATTC 401
|||||
106 GGGTGAACCTCCTGGAGGACATGTTGAAGACATCAATTCGACATATCTATACCT 165
|||||
400 GGGTGAACCTCCTGGAGGACATGTTGAAGACATCAATTCGACATATCTATACCT 341
|||||
166 GAATACAAAGAGCTTCTTCAAGAGTTCATAGACAGTATCCGCTCAGAGGCTATGGG 225
|||||
340 GAATACAAAGAGCTTCTTCAAGAGTTCATAGACAGTATCCGCTCAGAGGCTATGGG 281
|||||
226 AAATTCAGAGTGTTCCTCAACAGTCAATAGAACTCTGAAAAAATTTGGAGTGAAT 285
|||||
280 AAATTCAGAGTGTTCCTCAACAGTCAATAGAACTCTGAAAAAATTTGGAGTGAAT 221
|||||
286 ATGCATACAGTGTACAGACAGCATTTGGTGAATATGAAGATTAATTAATTCACCAAG 345
|||||
220 ATGCATACAGTGTACAGACAGCATTTGGTGAATATGAAGATTAATTAATTCACCAAG 161
|||||
346 CGTTGGCTCAGAGGCTACAGACTATGCGCAGAACTCATCTGTTGCTTGAAGCA 405
|||||
160 CGTTGGCTCAGAGGCTCAGAGCTATGCGCAGAACTCATCTGTTGCTTGAAGCA 101
|||||
406 CTTTCTTCTTGTG--TTGCTTTTATGTGGAACTGCTAGACAACCTGTTGAACCTCAA 462
|||||
100 CTTTCTTCTTGTGTTCTTTTATGTGGAACCTGCTAGACAACCTGTTGAACCTCAA 41
|||||
```

```

QY 463 TT 464
Db 40 AT 39

RESULT 8
A1936084/c 471 bp mRNA linear EST 02-SEP-1999
LOCUS w061d06.x1 NCI-CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2459819 3'
DEFINITION similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR. ; mRNA
sequence.
ACCESSION A1936084
VERSION A1936084.1 GI:5674954
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 471)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/dbrf/image/image.html
Seq primer: -40UP from Gbco
High quality sequence stop: 433.
FEATURES
source
location/Qualifiers
1..471
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2459819"
/clone_lib="NCI-CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT733 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 131 a 94 c 103 g 143 t
ORIGIN
Query Match 81.8%; Score 389.6; DB 9; Length 471;
Best Local Similarity 97.1%; Pred. No. 7.7e-101;
Matches 408; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 46 ATGAGCTGCTGATGCTCCTCATGCTGGCGCCCTCCTCTGCAGCTCTATGCAGATTCT 105
|||||
Db 471 ATGAAGCTGTGATGGCTCATGCTGGCGCCCTCCTCTGCAGCTCTATGCAGATTCT 412
|||||

QY 106 GGGTGCAGAACTCCGTGGAGCATGCTTGAAGACATCAATCCGACATATCTATCT 165
|||||
Db 411 GGCTGCAAACTCCGTGGAGCATGCTTGAAGACATCAATCCGACATATCTATCT 352
|||||

QY 166 GAATACAAAGAGCTTCTTCAAGATTATAGACAGTATGCCCTGCAGAGCTATGGG 225
|||||
Db 351 GAATACAAAGAGCTTCTTCAAGATTATAGACAGTATGCCCTGCAGAGCTATGGG 292
|||||

QY 226 AAATTCAGAGCTGTTCTTCAACCATCATAGAACTGTGAAAACTTTGACATGAT 285
|||||

```

```

Db 291 AAATTCAGAGCTGTTCTTCAACCATCATAGAACTGTGAAAACTTTGACATGATG 222
QY 286 ATGCATACAGTGTACGACACATTTGGTGTAATATGAGAGTAATTAACTTACCCAGG 345
|||||
Db 231 ATGCATACAGTGTACGACACATTTGGTGTAATATGAGAGTAATTAACTTACCCAGG 172
|||||

QY 346 CGTTGGCTCAGAGGCTACAGACTATGGCCAGAACATCATCTGTGATTGCTGAAGACCA 405
|||||
Db 171 CGTTGGCTCAGAGGCTACAGACTATGGCCAGAACATCATCTGTGATTGCTGAAGACCA 112
|||||

QY 406 CTTTCTTCTGTGCTGCTTTTATGTGGGAACCTGTGACAACTGTTGAACCTCAA 462
|||||
Db 111 CTTTCTTCTGTGCTGCTTTTATGTGGGAACCTGTGACAACTGTTGAACCTCAA 52
|||||

RESULT 9
AW779377/c 458 bp mRNA linear EST 12-MAY-2000
LOCUS hn79g08.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3034142 3'
DEFINITION similar to TR:075556 075556 MAMMAGLOBIN B PRECURSOR. ; mRNA
sequence.
ACCESSION AW779377
VERSION AW779377.1 GI:7793980
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 458)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gbco
High quality sequence stop: 458.
FEATURES
source
location/Qualifiers
1..458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3034142"
/clone_lib="NCI-CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clone IDs 1322376-1323911, 1456007-1456779, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "
BASE COUNT 132 a 94 c 99 g 132 t 1 others
ORIGIN
Query Match 81.7%; Score 389; DB 10; Length 458;
Best Local Similarity 96.7%; Pred. No. 1.1e-100;
Matches 408; Conservative 0; Mismatches 11; Indels 3; Gaps 1;

QY 46 ATGAGCTGCTGATGCTCCTCATGCTGGCGCCCTCCTCTGCAGCTCTATGCAGATTCT 105
|||||
Db 458 ATGAAGCTGTGATGGCTCATGCTGGCGCCCTCCTCTGCAGCTCTATGCAGATTCT 399
|||||

```


QY 106 GGCTGCAGAACTCCTGGAGCAGCATGTTGAAAGAACCATCAATTCGACATATCTATACCT 165
 |||||
 Db 398 GGGTGCAGAACTCCTGGAGCAGCATGTTGAAAGAACCATCAATTCGACATATCTATACCT 339
 QY 166 GAATACAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 225
 |||||
 Db 338 GAATACAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 279
 QY 226 AATTCAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 285
 |||||
 Db 278 AATTCAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 219
 QY 286 ATGCATACAGTGTACGACAGCATTTGGTGAATATGAGAGTAACTTACCTTACCCAGG 345
 |||||
 Db 218 ATGCATACAGTGTACGACAGCATTTGGTGAATATGAGAGTAACTTACCTTACCCAGG 159
 QY 346 CGTTGGCTCAGAGGCTACAGACTATGGCCAGAACTCATCTGTTGATTCCTGAAACCA 405
 |||||
 Db 158 CGTTGGCTCAGAGGCTACAGACTATGGCCAGAACTCATCTGTTGATTCCTGAAACCA 99
 QY 406 CTTTCTCTGTG---TTGCTTTTATGTGGGAGCTGTAGACACTGTGAAACCTCAA 462
 |||||
 Db 98 CTTTCTCTGTGTGTTCTTTTATGTGGGAGCTGTAGACACTGTGAAACCTCAA 39
 QY 463 TT 464
 |
 Db 38 AT 37

RESULT 10 458 bp mRNA linear EST 12-MAY-1999
 A1491987
 LOCUS A1491987.1 GI:4393001
 DEFINITION similar to SW:MAMG_HUMAN Q13296 MAMMAGLOBIN PRECURSOR. [1] ;, mRNA
 sequence.
 ACCESSION A1491987
 VERSION A1491987.1
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 458)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bdnp/image/image.html
 Insert Length: 577 Std Error: 0.00
 Seq primer: -40UP from Glibco
 High quality sequence stop: 399
 POLY-A-NO.

FEATURES
 source
 1..458
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2178381"
 /clone_lib="NCI-CGAP-U2"
 /tissue_type="moderately-differentiated endometrial
 adenocarcinoma, 3 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.85 kb. Life Technologies catalog #:

BASE COUNT 11539-012"
 132 a 93 c 101 g 132 t
 ORIGIN
 Query Match 80.9%; Score 385.2; DB 9; Length 458;
 Best Local Similarity 96.2%; Pred. No. 1.4e-99;
 Matches 406; Conservative 0; Mismatches 13; Indels 3; Gaps 1:
 QY 46 ATGACGCTGTGATGCTCTCATGCTGCGCGCCCTCTCTCGACGCTATGACAGATTCT 105
 |||||
 Db 458 ATGACGCTGTGATGCTCTCATGCTGCGCGCCCTCTCTCGACGCTATGACAGATTCT 399
 QY 106 GGGTGCAGAACTCCTGGAGCAGCATGTTGAAAGAACCATCAATTCGACATATCTATACCT 165
 |||||
 Db 398 GGGTGCAGAACTCCTGGAGCAGCATGTTGAAAGAACCATCAATTCGACATATCTATACCT 339
 QY 166 GAATACAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 225
 |||||
 Db 338 GAATACAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 279
 QY 226 AATTCAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 285
 |||||
 Db 278 AATTCAGAGCTTCTTCAAGAGTTCATAGACAGTGCATCCGCTGCAGAGGCTATGGGG 219
 QY 286 ATGCATACAGTGTACGACAGCATTTGGTGAATATGAGAGTAACTTACCTTACCCAGG 345
 |||||
 Db 218 ATGCATACAGTGTACGACAGCATTTGGTGAATATGAGAGTAACTTACCTTACCCAGG 159
 QY 346 CGTTGGCTCAGAGGCTACAGACTATGGCCAGAACTCATCTGTTGATTCCTGAAACCA 405
 |||||
 Db 158 CGTTGGCTCAGAGGCTACAGACTATGGCCAGAACTCATCTGTTGATTCCTGAAACCA 99
 QY 406 CTTTCTCTGTG---TTGCTTTTATGTGGGAGCTGTAGACACTGTGAAACCTCAA 462
 |||||
 Db 98 CTTTCTCTGTGTGTTCTTTTATGTGGGAGCTGTAGACACTGTGAAACCTCAA 39
 QY 463 TT 464
 |
 Db 38 AT 37

RESULT 11 396 bp mRNA linear EST 18-APR-1997
 AA297402
 LOCUS AA297402
 DEFINITION similar to steroid-binding protein, C3 chain, prostate, mRNA
 sequence.
 ACCESSION AA297402
 VERSION AA297402.1
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 396)
 AUTHORS Adams,M.D., Kervatage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
 ,C.J., Lee,M.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
 ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C., Clayton,R.A.,
 Cline,T.R., Cotton,M.D., Earle-Hughes,J., Flier,L.D., Fitzgerald,
 L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghegan,N.S., Glodex,A.,
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
 Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,
 Moreno-Palacios,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,U.L., Sauder,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
 Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
 Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
 ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
 Venter,J.C.
 TITLE Initial assessment of human gene diversity and expression patterns

LOCUS AA493295 435 bp mRNA linear EST 19-AUG-1997
DEFINITION ng97e06.s1 NCI-CGAP_Thy1 Homo sapiens cDNA clone IMAGE:942754
similar to TR:G1199596 G1199596 MAMMAGLOBIN.; mRNA sequence.
ACCESSION AA493295
VERSION AA493295.1 GI:2223136
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 435)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbip/image/image.html
Insert Length: 887 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 401.
FEATURES
source
1..435
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="942754"
/clone_lib="NCI-CGAP_Thy1"
/tissue_type="thyroid"
/lab_host="DH10B"
/ptc="Vector: PAMP10; mRNA made from invasive thyroid
tumor; cDNA made by oligo-dT priming. Non-directionally
cloned. Size-selected on agarose gel, average insert size
600 bp. Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 122 a 86 c 94 g 133 t
ORIGIN
Query Match 76.6%; Score 364.6; DB 9; Length 435;
Best Local Similarity 97.0%; Pred. No. 1.1e-93;
Matches 383; Conservative 0; Mismatches 9; Indels 3; Gaps 1;
QY 73 GCGGCCCTCCCTCGACACTGCTATGAGATTCTGGTCGCAAACTCCTGAGACATGCTT 132
|||||
DB 435 GCGGCCCTCCCTCGACACTGCTATGAGATTCTGGTCGCAAACTCCTGAGACATGCTT 376
QY 133 GAAAGACCTCAATTCGACATATCTATACCTGAATACAAGAGCTTCTTCAAGATTC 192
|||||
DB 375 GAAAGACCTCAATTCGACATATCTATACCTGAATACAAGAGCTTCTTCAAGATTC 316
QY 193 ATGAGAGTATGCGCGTGAAGAGGCTATGGGAAATTCAGAGAGTGTTCCTCAACAG 252
|||||
DB 315 ATGAGAGTATGCGCGTGAAGAGGCTATGGGAAATTCAGAGAGTGTTCCTCAACAG 256
QY 253 TCACATGAGACTGTGAAGAACTTTGGACTGATGATGATGATGATGATGATGATGATG 312
|||||
DB 255 TCACATGAGACTGTGAAGAACTTTGGACTGATGATGATGATGATGATGATGATGATG 196
QY 313 TGTATATGAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 372
|||||
DB 195 TGTATATGAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 136
QY 373 GGCAGAGTATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 429
|||||
DB 135 GGCAGAGTATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 76
QY 430 GTGGAGACTGCTAGACCACTGTGAAACCTCAATT 464

Db 75 GTGGAACCTCTAGACAACTGTGAAACCTCAATT 41
RESULT 14
LOCUS BE044895
DEFINITION BE044895 365 bp mRNA linear EST 08-JUN-2000
hno8d03.x1 NCI-CGAP_Thy6 Homo sapiens cDNA clone IMAGE:3021509 3'
similar to TR:O75556 O75556 MAMMAGLOBIN B PRECURSOR.; mRNA
sequence.
ACCESSION BE044895
VERSION BE044895.1 GI:8361948
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 365)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike
Emmert-Buck M.D., Ph.D., Vlado Knezevic M.D.
CDNA Library Preparation: Krizman Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium/LLNL
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Possible reversed clone: similarity on wrong strand
possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 364.
FEATURES
source
1..365
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3021509"
/clone_lib="NCI-CGAP_Thy6"
/tissue_type="normal epithelium"
/lab_host="DH10B"
/Note="Organ: thyroid; Vector: PAMP10; mRNA made from
normal thyroid epithelium; cDNA made by oligo-dT priming.
Non-directionally cloned into UDG sites. Size-selected on
agarose gel, average insert size 500 bp. Primary library.
CDNA Library Preparation: David B. Krizman, Ph.D.
REFERENCE: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 103 a 85 c 85 g 92 t
ORIGIN
Query Match 75.0%; Score 357; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.6e-91;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 50 AGCTGCTGATGATGCTTCATGCTGCGCGCCCTCCTCTGCTGACATGCTATGAGATTCTGGCT 109
|||||
DB 9 AGCTGCTGATGATGCTTCATGCTGCGCGCCCTCCTCTGCTGACATGCTATGAGATTCTGGCT 68
QY 110 GCAAACTCCTGAGAGACATGCTGAAAAAGACATTAATTCGACATATCTATACCTGAAT 169
|||||
DB 69 GCAAACTCCTGAGAGACATGCTGAAAAAGACATTAATTCGACATATCTATACCTGAAT 128
QY 170 ACAAGAGCTTCTTCAAGAGTATATACAGAGTATGATGATGATGATGATGATGATGATG 229
|||||
DB 129 ACAAGAGCTTCTTCAAGAGTATATACAGAGTATGATGATGATGATGATGATGATGATG 188
QY 230 TCAAGAGTGTTCCTCAACACAGTACATAGAACTGTGAAAAAATTTTGAAGTATGATGTC 289
|||||
DB 189 TCAAGAGTGTTCCTCAACACAGTACATAGAACTGTGAAAAAATTTTGAAGTATGATGTC 248

QY	290	ATACAGTCTACACAGCACTTTGGTGAATATGCAAGTAATTAACCTTACCAACGCGTT	349
Db	249	ATACAGTCTACACAGCACTTTGGTGAATATGCAAGTAATTAACCTTACCAACGCGTT	308
QY	350	TGGCTCAGAGGGCTACAGACTATGGCCGCAACATCATCTGTGTGATTGCTAGTAACAC	406
Db	309	TGGCTCAGAGGGCTACAGACTATGGCCGCAACATCATCTGTGTGATTGCTAGTAACAC	365
RESULT 15			
LOCUS	BE044893	351 bp	mrna
DEFINITION	hnb8c04.x1 NCI-CGAP-Thy6 Homo sapiens cDNA clone IMAGE:3021510 3'	linear	EST 08-JUN-2000
ACCESSION	BE044893		
VERSION	BE044893.1	GI:8361946	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 351)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@b5-remail.nih.gov		
	Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike Emmert-Buck M.D., Ph.D., Vlado Knezevic M.D.		
	cDNA Library Preparation: Krizman Laboratory		
	cDNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov		
	Info@image.llnl.gov		
	Possible reversed clone: similarity on wrong strand		
	Possible reversed clone: polyt not found		
	Seq primer: -40up from Glbco		
	High quality sequence stop: 351.		
FEATURES			
Source	Location/Qualifiers		
	1..351		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3021510"		
	/clone_lib="NCI-CGAP-Thy6"		
	/tissue_type="normal epithelium"		
	/lab_host="DH10B"		
	/note="Organ: thyroid; Vector: pAMP10; mRNA made from normal thyroid epithelium, cDNA made by oligo-dT priming. Non-directionally cloned into udc sites. Size-selected on agarose gel, average insert size 500 bp. Primary library. cDNA library preparation: David B. Krizman, Ph.D. REFERENCE: Krizman et al. (1996) Cancer Research 56:5380-5383."		
BASE COUNT	98 a 81 c 83 g 89 t		
ORIGIN			
Query Match	72.1%	Score 343;	DB 10; Length 351;
Best Local Similarity	100.0%;	Prod. No. 1.6e-87;	
Matches 343;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	50	AGCTCTGATGATGCTCTCAAGCTGAGGGCCGCTCCCTGACAGCTATGACAGATTCGGCT	109
Db	9	AGCTCTGATGATGCTCTCAAGCTGAGGGCCGCTCTCTCTGACAGCTATGACAGATTCGGCT	68
QY	110	GCAAACTCTGTGAGACATGTTGTAAGAACACATCAATTCGACATATCTATACCTGAAT	169
Db	69	GCAAACTCTGTGAGACATGTTGTAAGAACACATCAATTCGACATATCTATACCTGAAT	128
QY	170	ACAAAGAGCTTCTTCAAGAGTTCAATGACAGTGAATGCCGTGACAGAGCTATGGGAAAT	229

Db 129 ACAAGAGCTTCTTCAGAGTTCATTAGACAGTGTGATGGCGCTGCAGAGGCTATGTGGCAAT 188
|
|
QY 230 TCAAGCAGTGTTCCTCAACCACTGACATAGAACTGTGAAAACCTTTGGACTGATGTC 289
|
Db 189 TCAAGCAGTGTTCCTCAACCACTGACATAGAACTGTGAAAACCTTTGGACTGATGTC 248
|
QY 290 ATACAGTGTACGACAGCATTTGGTGTAAATGTGAAGACTAATTAACCTTACCAGAGCGTT 349
|
Db 249 ATACAGTGTACGACAGCATTTGGTGTAAATGTGAAGACTAATTAACCTTACCAGAGCGTT 308
|
QY 350 TGGCTCAGAGGGGCTACAGACTATGGCCAGAACTCATCTGTTGA 392
|
Db 309 TGGCTCAGAGGGGCTACAGACTATGGCCAGAACTCATCTGTTGA 351

Search completed: January 25, 2003, 13:18:08
Job time : 2260 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 10:29:19 ; Search time 3276 Seconds

(without alignments)
4228.613 Million cell updates/sec

Title: US-09-806-302A-1

Perfect score: 476
Sequence: 1 acgagctgcacgcacgact.....ccatcattcattca 476

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GeneBank: 1: gb_da: 2: gb_hvg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_ph: 24: em_pl: 25: em_ro: 26: em_sts: 27: em_un: 28: em_vl: 30: em_hvg_hum: 31: em_hvg_in: 32: em_hvg_other: 33: em_hvg_mus: 34: em_hvg_pln: 35: em_hvg_rtd: 36: em_hvg_mam: 37: em_hvg_vrt: 38: em_sy: 39: em_hvg_hum: 40: em_hvg_mus: 41: em_hvg_other:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	476	100.0	476	6	ARI42399
2	476	100.0	476	6	ARI82545
3	434.2	91.2	496	6	HSN224173
4	431.6	90.7	517	6	AX013093
5	431.6	90.7	517	9	AF071219
6	407.4	85.6	491	6	AX335698
7	223.8	47.0	495	6	ARI72283
8	223.8	47.0	503	6	AR080285
9	223.8	47.0	503	6	AR095403
10	223.8	47.0	503	6	AX213242
11	223.8	47.0	503	6	165735
12	223.8	47.0	503	9	HSU33147
13	223.8	47.0	535	6	AR207542
14	222.2	46.7	503	6	AX283024
15	211.6	44.5	429	6	AX213264
16	210	44.1	429	6	AX213261
17	208.4	43.8	429	6	AX213258
18	208.4	43.8	429	6	AX213259
19	208.4	43.8	429	6	AX213260
20	208.4	43.8	429	6	AX213263
21	205.6	43.2	388	6	AX213235
22	202.4	42.5	388	6	AX213234
23	202.4	42.5	388	6	AX213236
24	202.4	42.5	388	6	AX213237
25	202.4	42.5	388	6	AX213238
26	202.4	42.5	388	6	AX213239
27	202.4	42.5	388	6	AX213240
28	202.4	42.5	388	6	AX213241
29	201.2	42.3	403	6	AR080288
30	201.2	42.3	403	6	AR095406
31	201.2	42.3	403	6	165738
32	195.2	41.0	467	6	AX156186
33	191.8	40.3	17281	2	AC074200
34	191.8	40.3	17348	9	AP003306
35	191.8	40.3	178770	2	AC090725
36	191.8	40.3	181140	9	AP002793
37	188.4	39.6	420	6	AX213265
38	188.4	39.6	420	6	AX213270
39	186.8	39.2	420	6	AX213262
40	182.4	38.3	379	6	AX213233
41	172	36.1	506	4	AF308618
42	169.2	35.5	461	4	AF308620
43	169	35.5	486	6	AX358804
44	169	35.5	486	6	AX362297
45	160.8	33.8	452	4	AF308619

ALIGNMENTS

RESULT 1
LOCUS ARI42399
DEFINITION Sequence 5 from patent US 6174992.
ACCESSION ARI42399
VERSION ARI42399.1 GI:15102699
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 476)
AUTHORS N.J., Yu, G.-L. and Gentz, R.
TITLE Human endometrial specific steroid-binding factor I, II and III
JOURNAL Patent: US 6174992-A 5 16-JAN-2001;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

Dd	61	CATGCTGGCGGCCCTCCCTCGACAGCGTATGACAGATTCTGGTCGAACACTCCTGGAGAGA	120
Oy	126	CATGTTGAAAAAGACCATTCAATTCGGACATATCTATACCTGAATACAAAGCTTTCTTCA	185
Dd	121	CATGTTGAAAAAGACCATTCAATTCGGACATATCTATACCTGAATACAAAGAGCTTTCTTCA	180
Oy	186	AGAGTTCATTAAGACGTATGATGGCGGCTGCAGAGAGCTTTGGSGAATTCACAAGAGTGTTTCT	245
Dd	181	AGAGTTCATTAAGACGTATGATGGCGGCTGCAGAGAGCTTTGGSGAATTCACAAGAGTGTTTCT	240
Oy	246	CAACCACTCACATAGAACCTCTGAAAAACTTTGGAGCTGATGATCATACAGTAGTACAGACAG	305
Dd	241	CAACCACTCACATAGAACCTCTGAAAAACTTTGGAGCTGATGATCATACAGTAGTACAGACAG	300
Oy	306	CATTGGTGTATAATGAGAGTAATTAACCTTACCAGAGGCTTTGGCTCAGAGGCTAC	365
Dd	301	CATTGGTGTATAATGAGAGTAATTAACCTTACCAGAGGCTTTGGCTCAGAGGCTAC	360
Oy	366	AGACATATGGCCGAGAACCTCATCTGTTGATTGGTAGAACAACACTTT--CTTCTTGGTGTCT	423
Dd	361	AGACATATGGCCGAGAACCTCATCTGTTGATTGGTAGAACAACACTTTCTTTTGTGTCT	420
Oy	424	TTTTTATGTGGCACTGCTAGACACTGTTGAACCTCAATT	464
Dd	421	TTTTTATGTGGCAACCTGCTAGACACTGTTGAACCTCAAAAT	461
<hr/>			
RESULT 4			
AXO13093			
LOCUS	AXO13093	517 bp	DNA linear PAT 07-SEP-2000
DEFINITION	Sequence 33 from Patent WO9594461.		
ACCESSION	AXO13093		
VERSION	AXO13093.1 GI:10040259		
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Euharvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 517) Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pilarsky,C. Human nucleic acid sequences of endometrium tumour tissue Patent: WO 9594461-A 33 28-OCT-1999. SCHMITT AMIN (D); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN (DE); PILARSKY CHRISTIAN (DE)		
AUTHORS			
TITLE			
JOURNAL			
FEATURES	source location/Qualifiers 1..517 /organism="Homo sapiens" BASE COUNT 156 a 121 c 106 g 134 t ORIGIN		
Query Match	90.7% Score 431.6; DB 6; Length 517;		
Best Local Similarity	97.4%; Pred. No. 1.5e-117;		
Matches 450; Conservative	0; Mismatches 9; Indels 3; Gaps 1;		
Oy	6	CTGGCAGCAGCAGCTGAAGAACAGACAGAGCCGCTCGCATGAGCTGATGATGCTCT	65
Dd	1	CTGGCAGCAGCAGCTGAAGAACAGACAGAGCCGCTCGCATGAGCTGATGATGCTCT	60
Oy	66	CATGCTGGGGGCGCCCTCCTCTGCTGACTGCTATGACAGATTTCTGGCTGCAACCTCTGGAGGA	125
Dd	61	CATGCTGGGGGCGCCCTCCTCTGCTGACTGCTATGACAGATTTCTGGCTGCAACCTCTGGAGGA	120
Oy	126	CATGTTGAAAAAGACCATTCAATTCGGACATATCTATACCTGAATACAAAGAGCTTTCTTCA	185
Dd	121	CATGTTGAAAAAGACCATTCAATTCGGACATATCTATACCTGAATACAAAGAGCTTTCTTCA	180
Oy	186	ACAGTTTCATAGACAGTATGATCCGCTGCAGAGGCTATGGGAAAATTCACAAGCACTGTTTCT	245
Dd	181	ACAGTTTCATAGACAGTATGATCCGCTGCAGAGGCTATGGGAAAATTCACAAGCACTGTTTCT	240

OY	246	CACAGCAGCACATACAACCTCGAAAAAACCCTTGCGATGTAAGTCACTACAGACG	305
Db	241	CACACAGTCACATAAACCTCGAAAAAACCCTTGCGATGTAAGTCACTACAGACG	300
OY	306	CATTGGCTGTATATGAAGAATAATTAACCTTACCCAAAGCGTTGGCTCAGAGGCCTAC	365
Db	301	CATTGGGTATATATGAAGAATAATTAACCTTACCCAAAGCGTTGGCTCAGAGGCCTAC	360
OY	366	AGACTATGGCCAGAACTCATCTGTTGATTGCTAGAAACCACTTCTTCTGTG--TTGC	422
Db	361	AGACTATGGCCAGAACTCATCTGTTGATTGCTAGAAACCACTTCTTCTGTGTGTC	420
OY	423	TTTTATGTGGGAACGTGAGACCACTGTGTAACCTCAATT	464
Db	421	TTTTATGTGGGAACGTGAGACCACTGTGTAACCTCAATT	462
RESULT 5	AF071219		
LOCUS	AF071219	517 bp	mRNA linear PRI 26-NOV-1998
DEFINITION	Homo sapiens mammaglobin B precursor, mRNA, complete cds.		
ACCESSION	AF071219		
VERSION	AF071219.1	GI:3288867	
KEYWORDS	.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 517) Becker,R.M., Darow,C., Zimonjic,D.B., Popescu,N.C., Watson,M.A. and Fleming,T.P. Identification of mammaglobin B, a novel member of the uteroglobin gene family Genomics 54 (1), 70-78 (1998)		
JOURNAL	JOURNAL	99026127	
MEDLINE	PUBMED	9806831	
REFERENCE	2 (bases 1 to 517)		
AUTHORS	Becker,R.M., Darow,C.M., Watson,M.A. and Fleming,T.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JUN-1998) Ophthalmology, Washington University School of Medicine, 660 S. Euclid, St. Louis, MO 63110, USA		
FEATURES	Location/Qualifiers		
source	1..517	/organism="Homo sapiens"	
	/db_xref="taxon:9606"		
	/chromosome="11"		
	/map="11q13"		
CDS	65..352	/note="member of the uteroglobin family"	
	/codon_start=1		
	/product="mammaglobin B precursor"		
	/protein_id="AAC79996.1"		
	/db_xref="gi:3288868"		
	/translation="MKILMTVLMAALLHCVAADSGCKLLEDVMEKTIINSIDISEYREK LLOEIDSDDAAPAMGKKQGFLLNOSHRTTLNPFGLMHHTVYDLSIMCMKSN"		
BASE COUNT	144 a 129 c 105 g 139 t		
ORIGIN			
Query Match	90.7%; Score 431.6; DB 9; Length 517;		
Best Local Similarity	97.4%; Pred. NO.1.5e-117;		
Matches 450; Conservative	0; Mismatches 9; Indels 3; Gaps 1;		
OY	6	CTGCCACGACGACATGAGACAGACAGACGAGCGCGCTCGCATGAAAGCTGTATGCTCT	65
Db	25	CTGCCACGACGACATGAGACAGACAGACGAGCGCGCTCGCATGAAAGCTGTATGCTCT	84
OY	66	CATGCTGGGGGGCCCTCTCTCTGACATGCGATGACAGATTCTGGCTGCAAACTCTGGAGGA	125
Db	85	CATGCTGGGGGGCCCTCTCTCTGACATGCGATGACAGATTCTGGCTGCAAACTCTGGAGGA	144
OY	126	CATGCTTGAAGAACCATCAATTCGCATATCTATACCTGAATCAAGAAAGACTTCTTCA	185

[illegible]

BASE COUNT	146 a	118 c	97 g	142 t
ORIGIN	/organism="unknown"			
Query Match	47.0%: Score 223.8; DB 6; Length 503;			
Best Local Similarity	71.4%: Pred. No. 1.4e-55;			
Matches 325; Conservative	0; Mismatches 122; Indels 8; Gaps 2;			
QY	7 TGCACGACGACGCTAAACAGACAGCAGCGCCGCTGCGCATGAAAGCTGCTGATGTCCTC 66			
DB	22 TGCACCGCCGCGACTAACACGACGACGACAGCCCTACCATGAAAGTCTCTATGTCCTC 81			
QY	67 ATGCTGGCGGCCCTCTCTGTCGCTATGACGATTCGTGGCTGCGAACTCCTGAGAGAC 126			
DB	82 ATGCTGGCGGCCCTCTCTGTCGCTATGACGAGCTCTGGGCTGCCCTTATTTGAGAAAT 141			
QY	127 ATGCTTGAAGAACCATCATTTCCGACATATCTTAACTGAATFACAAAGAGCTTCTTAA 186			
DB	142 GGTATTTTCCAAACATCAATTCACAAAGTCTTAGAGCTGAATACMAAGAACTTCTTCAA 201			
QY	187 GAGTTTCATGACAGTATGCGCCGCTGACAGGCTATGGGAAATTCAGCAGTGTTCCTC 246			
DB	202 GAGTTTCATGACAGCAGATGCGCTCTCAAAATGCCATAGATGAATTAAGAAATGTTTCTT 261			
QY	247 AACCAATCATAGACTCTGMAAACTTTGGACATGATGACATACAGTACGACAGC 306			
DB	262 AACCAAGCGAGATGAAAGACTCTGACAGATGTGGAGGTGTTATGATTAATATGACAGC 321			
QY	307 ATTTGGTGAATATGACAGCTAATTAACCTTTACCAAGGCGTTGGCTCAGAGGCTTACA 366			
DB	322 AGCTCTTTTGATTT-----ATTTTAACTTCTTGCMAAACCTTTGGCTCAGAGAACTCA 375			
QY	367 GACTATGGCCACAACTCATCTGTTATGCTATGCTAGAAC--CACTTCTCTGTTGTCCT 424			
DB	376 GCGTATGAGGAGAAACCACTAGAGATGTCGTCAAAACCAACACACTTCTCTTCTTATGTC 435			
QY	425 TTTATGTGGGAACCTGCTAGACAACTGTTGAAACCT 459			
DB	436 TTTTACTACAACTFACAGCAATGTTGAAACCT 470			
RESULT 10				
LOCUS	AX213242	503 bp	DNA	Linear
DEFINITION	Sequence 27 from Patent WO0158947.			
ACCESSION	AX213242			
VERSION	AX213242.1	GI:15524172		
KEYWORDS	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Carter,D., Vedvick,T.S., Vallieve-Douglas,J., Houghton,R.L. and Dillon,D.C.			
TITLE	Lipophilin complexes for use in cancer diagnosis and therapy			
JOURNAL	Patent: WO 0158947-A 27 16-AUG-2001;			
FEATURES	CORLXA CORPORATION (US)			
source	Location/Qualifiers			
	1..503			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
BASE COUNT	146 a	118 c	97 g	142 t
ORIGIN				
Query Match	47.0%: Score 223.8; DB 6; Length 503;			
Best Local Similarity	71.4%: Pred. No. 1.4e-55;			
Matches 325; Conservative	0; Mismatches 122; Indels 8; Gaps 2;			
QY	7 TGCACGACGACGCTAAACAGACAGCAGCGCCGCTGCGCATGAAAGCTGCTGATGTCCTC 66			
DB	22 TGCACCGCCGCGACTAACACGACGACGACAGCCCTACCATGAAAGTCTCTATGTCCTC 81			

OY	ATCCGCGGGCCCTCCTCCTGCACTGCTATGCGAATTTGGGTGCMAAATCCTCGAGAGAC	126
Db	ATGCGCGGGGCCCTCTCCAGACACTGCTACAGAGGCTCTGGCTGCCCTTATTTGGACAAT	141
OY	ATGCGTTGAAAAGACCATCAATTCGACATATCTATATACCTGAAATACAAAGAGCTTCTTCAA	186
Db	GTGATTTCCAAAGACAAATCAATCCACAAAGTGTCTAAGACTGAAATACAAAGAACTTCTTCAA	201
OY	GAGTTCTATAGACAGTGGATGCCCTCGACAGAGCTATGGGGAATTCAGCAGTGTCTTC	246
Db	GAGTTCAATAGACAGCAATGCGCCACTGCAAAATGCGAATAGTAATTCGAAGAAATGTTTCTT	261
OY	AACCAAGTCAATAGAAAGCTGTGAAAACATTTGGACTGATGATGACATACAGTGTACACAGC	306
Db	AACCAAAACGATGAAGAACTCTGAGCAATGTTGAGGTGTTATATGCAATTAATATATGACAGC	321
OY	ATTTGGGTATATATGAAAGATTAATTAACCTTTACCCAAAGGCGTTTGGCTCAGAGGGCTACA	366
Db	ACTCTTTTGATATTT-----ATTTTAACCTTTCTGCAAGACCTTTGGCTCAGCAAGACTGCA	375
OY	GACTATGGCCGAGACTCATGTGTGATTTGCTAGAAAC--CACTTTCTTCTGTGTGCTT	424
Db	GGGTATGGTGAGAAAACCAACTACGGAATTTGTCGAAACCAACCACTCTCTTCTTATGTCT	435
OY	TTTATGTGGGAACCTGCTAGACAACAGTTGAAACCT	459
Db	TTTACTACAAACTACAAAGCAATTTGTAACCT	470

RESULT	11				
LOCUS	165735				
DEFINITION	Sequence 1 from patent US 5668267.	503 bp	DNA		Linear
ACCESSION	165735				PAT 07-OCT-1997
VERSION	165735.1	GI:2482305			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 503)				
TITLE	Watson,M.A. and Fleming,T.P. Polynucleotides encoding mammaploblin, a mammary-specific breast cancer protein				
JOURNAL	Patent: US 5668267-A 1 16-SEP-1997;				
FEATURES	Location/Qualifiers				
SOURCE	1..503				
	/organism="unknown"				
BASE COUNT	146 a	118 c	97 g	142 t	
ORIGIN					

Query Match	47.0%	Score 223.8	DB 6	Length 503
Best Local Similarity	71.4%	Pred. No. 1.4e-55		
Matches 325	Conservative 0	Mismatches 122	Indels 8	Gaps 2

Qy	7	TGGCAGGACAGCACTGAACACAGACAGCAGCGCCCTGGCATCAACCTGATGGTCTCG	66
Dy	22	TGGCAGCGCCGACACTGAACACGACGACGACAGCAGCCTCACCATTGAATTTGCTATGGTCTCG	81
Qy	67	ATGCTGGGGGGCCCTCCTCCTGCACTGCTATGTCAGATTCTGGCTGCAAACTCCTGGAGGAC	126
Dy	82	ATGCTGGGGGGCCCTCCTCCTGCACTGCTATGTCAGAGGCTCTGGCTGGCCCTTATTTGGAGAT	141
Qy	127	ATGGTTGAAAAGACCATCAATTCGACACATATCTATACCTGAATACAAAAGAGCTTCTCAA	186
Dy	142	GTGATTTTCCAGACCAATCAATCAATCAACAGACTGTCTTAAGACTGAAATACAAAAGAACTTTCTCAA	201
Qy	187	GAGTTTCATAGACAGTGTATGCCGCTGCAGAGGCTATGGGGAATTCACAGCAGTGTTCCTCG	246
Dy	202	GAGTTTCATAGACAGCAATGTGCCCACTACAAATGCCATATGAAATTAAGAAGAAATGTTTCTT	261
Qy	247	AACCACTCACTATAGAACTCTGAAAAAATTTTGGACTGATGATGCAATACAGTGTACACAGC	306
Dy	262	AACCAACAGGATGAACCTGTGAGCAATGTTTGAGGGTATTATCAATTTAATATATATACAGC	321

[illegible]

RESULT 12			
LOCUS	HSU33147	503 bp	linear
DEFINITION	Human mammapaglobin mRNA, complete cds.		
ACCESSION	U33147		
VERSION	U33147.1	GI:119595	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 503)	Watson, M.A. and Fleming, T.P.	Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer	Cancer Res. 56 (4), 860-865 (1996)

JOURNAL MEDLINE
PUBMED
REFERENCE
TITLE
AUTHORS
JOURNAL

Cancer Res. 56 (4), 860-865 (1996)
96232698
8631025
2 (bases 1 to 503)
Watson, M.A.
Direct Submission
Submitted (03-AUG-1995) Mark A. Watson, Washington Univ. School of
Medicine, Dept. Ophthalmology and Visual Sciences, Dept. Genetics,
Div. Laboratory Medicine, Box 8118, 660 S. Euclid Avenue, St.
Louis, MO 63110, USA

```

FEATURES
source
    location/Qualifiers
    1..503
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /tissue_type="breast"
    61..342
    /note="Similar to uteroglobin: SwissProt Accession Number
    P02775 and rat prostatic steroid binding protein subunit
    C3: SwissProt Accession Number P02780"
    CDS

```

```

/codon_start=1
/product="mammaglobin"
/protein_id="AAC50375.1"
/db_xref="GI:1199596"
/translation="MKILMVLMLAALSHOCHYAGSGCPDENIVSKTINPOVSNEYKE
LLOEFIDNATTNALIDELKECFLNQTDETLSNVEVFQLLYDSDLCLF"

```

[illegible]

[illegible]

RESULT	13				
LOCUS	AR207542	535 bp	DNA	linear	PAT 20-JUN-2002
DEFINITION	AR207542	Sequence 1 from patent US 63/9671.			
ACCESSION	AR207542				
VERSION	AR207542.1	GI:21507322			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 535)				
AUTHORS	Colpitts,T.L.				
TITLE	Reagents and methods useful for detecting diseases of the breast				
JOURNAL	Patent: US 63/9671-A 1 30-Apr-2002;				
FEATURES	Location/Qualifiers				
source	1..535				
	/organism="unknown"				
BASE COUNT	153 a 127 c 105 g 150 t				
ORIGIN					

Query Match	47.0%:	Score 223.8:	DB 6:	Length 533:
Best Local Similarity	71.4%:	Pred. No. 1.5e-55:		
Matches 325:	Conservative	0:	Mismatches 122:	Indels 8:
			Gaps	2:
QY 7	TGCCACGCACAGCTGAACACAGACAGACGCCGCTCGGCATGAAGTCTCTGATGGTCCTC	66		
Db 43	TGCCACCGCGACCTGAACACCGAGACGACGCTCACCATGAATTCCTATATGGTCCTC	102		
QY 67	ATGCTGGGGGCCCTCTCTCTGACTGCTATGACAGATTGTGGCTGCAAACTCCTGGAGAC	126		
Db 103	ATGCGGGGGCCCTCTCCACACACTGCATACGAGGCTCTGGCTGCCCTCTATTGGAGAT	162		
QY 127	ATGGTTGAAGAACCATCAATTCGCATATTCATATACCTGAAATACAAAGAGCTTTGCAA	186		
Db 163	GTGATTTCCAAAGACATCAATTCACAAAGTGTCTAAGATCTGAATACAAAGACTTTTCAA	222		
QY 187	GAGTTCATAGACAGTGAATGCCGCTGCAGAGGCTATGGGAAATTCAGACAGTGTTCCTC	246		
Db 223	GAGTTCATAGACAGCAAAATGCCACTCAAAATGCCATAGATGAATGAAGAATGTTTCTT	282		
QY 247	AACCGTCACATGAAACCTCTGAAAACCTTTGGAGCTGATGATGCACATACAGTACGACGC	306		
Db 283	AACCAACGAGTGAAGACCTCTAGCATGTGGTGTGTTATGACATTTAATATATATACAGC	342		
QY 307	ATTGGTGTAAATATGAAGAGTAATTAATTAACCAAGGCGTTGGCTCAGAGGGCTACA	366		
Db 343	AGCTTTTGATTT-----ATTTAACCTTCGTCAAGACCTTTGGCTCACGAACTGCA	396		
QY 367	GACTATGGCCGAACATCATCTGTTATATGCTGAAGAAC--CACTTCTCTGTGTGGCTT	424		
Db 397	GGATTTGTGTGAACCAACGCTTACGGATGTGCTCAACACACACACTCTCTTCTTATGTCT	456		

Oy 425 TTATGTGGAACTGCTAGACAACGTGTTGAAACCT 459
||| | |||| | ||||| ||||| |||||
Db 457 TTTTACTACAACACTACAAGACACTGTTGAAACCT 491

RESULT_14			
AX283024	AX283024	503 bp	DNA
LOCUS	Sequence 73 from Patent WO01/5171.		linear
DEFINITION			
ACCESSION	AX283024		
VERSION	AX283024.1 GI:16609951		

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 Houghton,R.L., Dillon,D.C., Molesh,D.A., Xu,J., Zehentner,B. and Persing,D.H.
TITLE	Methods, compositions and kits for the detection and monitoring of breast cancer
JOURNAL	Patent: WO 0175171-A 73 11-OCT-2001; CORIXA CORPORATION (US)
FEATURES	Location/Qualifiers
source	1. 503 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	145 a 119 c 97 g 142 t
ORIGIN	

Query Match	46.7%	Score 223.2	DB 6	Length 503
Best Local Similarity	71.2%	Pred. No. 4.3e-55		
Matches 324	Conservative 0	Mismatches 123	Indels 8	Gaps 2

Oy	7	TGCCACGACACACATGGAACACACAGACAGGAGGCGCCTGCACATGGAAGGCGTGAATGTC	66
Dd	22	TGCCACCGCCGACACTGAAACACCGACGACGACGCTCACCATGAAGTTGCTGATGTC	81
Oy	67	ATGCTGGGCGCCCTCTCTGCTGACACTGCTATGACAGATTCTGGCTGCAAACTCCTGGAGGAC	126
Dd	82	ATGCTGGGCGCCCTCTCTGCTGACACTGCTATGACAGATTCTGGCTGCAAACTCCTGGAGGAA	141
Oy	127	ATGCTGAAAGACACCTTCATCTCCGACATATCTTACTCTATATTCAAAGAGCTTTTCAA	186
Dd	142	GTGATTTCCAAGAACAAATCAATCCCAAGTGTCTTAAGACTAAATACAAAGAACTTCCTCAA	201
Oy	187	GAGTTCATAGACAGTGAATCGCGCTGCAGAGGCTGTGGGAAATTCAGACAGTGTTCCTC	246
Dd	202	CAGTTCATAGACAGCAATGCGCAGTCTACAAATGCCATATGATGATTTGAAGGAATGTTTCTT	261
Oy	247	AACGAGTCACATAGAACTCTGAAAAACTTTGGACTGATGATGATACATGATGACAGAC	306
Dd	262	AACCAAAAGGATGGAACCTCTGCAAAATTTGGAGGTGTTCTGCAATTAATATATGACAG	321
Oy	307	ATTGCTGTAAATATGAAGAGATTAATTAACCTTACCAGAGGGTTGGCTCAGAGGGCTACA	366
Dd	322	AGCTTTTGTGATTT-----ATTTAACTTTCTGCAAGACCTTTGGCTCACAACTGGCA	375
Oy	367	GACTATGGCCAGAACTCATCTGTTGATTTGCTAGAAAC--CACTTCTTCTTGTGTTGCTT	424
Dd	376	GGGTATGGTGAAGAAACCACTACAGGATTTGCTGCAAAACACACCTTCTCTTTATGTCT	435
Oy	425	TTTATGTGGGAATGCTAGACAACTGTTGAAACCT	459
Dd	436	TTTTACTACAAACTACAAAGAAATTTTGGAAACCT	470

RESULT 15	AX213264	LOCUS	AX213264	429 bp	DNA	linear	PAT 06-SEP-2001
DEFINITION	Sequence	49	from Patent WO0158947.				
ACCESSION	AX213264						
VERSION	AX213264.1	GI:15524179					

KEYWORDS	human.
SOURCE	
ORGANISM	Homo sapiens
REFERENCE	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 429)
AUTHORS	Carter,D., Vedvick,T.S., Vallée-Douglas,J., Houghcon,R.L. and Dillon,D.C.
TITLE	Lipophilin complexes for use in cancer diagnosis and therapy
JOURNAL	Patent: WO 0158947-A 49 16-AUG-2001; CORIXA CORPORATION (US)
FEATURES	Location/Qualifiers
source	1..429
BASE COUNT	/organism="Homo sapiens" /db_xref="taxon:9606"
ORIGIN	121 a 106 c 87 g 115 t
Query Match	44.5% Score 211.6; DB 6; Length 429;
Best Local Similarity	72.5%: Pred. No. 6,1e-52;
Matches 290: Conservative	0: Mismatches 104; Indels 6; Gaps 1;
QY	7 TCCACGCGACGACTAACAACAGACAGCAAGCCGCGCCATCAAGCTGCTGATGGTCCCTC 66
Db	19 TCGCACCCGCGACTAACAACGACGACGACGACCTCACCANTGAAATGGTGTGCTCTC 78
QY	67 ATCGTGGCGCGCCCTCTCTCGCACTGCTATGCGATTCGTGGCTGCMAACTCCTGGAGGAC 126
Db	79 ATGCTGGCGCGCCCTCTCCAGCAGCTGTACGCGAGGCTTGCGCTTATTGAGAAAT 138
QY	127 ATGCTGAAAAGACCATCAATTCGCACTATCTTATACCTGAAATACAAAGACTTTCTCA 186
Db	139 GTGATTTTCCAAAGACATCAATCCACAAGTGTCTAAGACGATACAAAGAACTTCTTCA 198
QY	187 GAGTTCATAGACAGTATGCGCTCGAGAGGTATGGGAAATTCAGACAGTCTTCCGC 246
Db	199 GAGTTCATAGACAGCAATATGCGCACTCAAAATGCGATAGATGAATTAAGGAATGTTTTCTT 258
QY	247 AACCACTACATAGAACTCTGAAAACCTTTGGACTGATGATGCATATCAAGTATGACAGC 306
Db	259 AACCAACGAGATGAAAGACTCTGAGCAATGTGAGTGCTTATATGCAATTAATATGACAGC 318
QY	307 ATTGGGTAAATATGAAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 366
Db	319 AGCTTTTGTGATTT-----ATTTTAACTTTCTGCAAGACCTTTGCTCAGCAAGACTGCA 372
QY	367 GACTATGGCCAGAACTCATCTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 406
Db	373 GGTATGGTGAAGAAACCAACTAGGATTTGATTAATTAATTAATTAATTAATTAATTAATTA 412

Search completed: January 25, 2003, 12:40:17
Job time : 3279 secs